THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 61, PART 1 1 July 1948 No. 343

The Atomic Heats of Palladium, Sodium and Mercury at Low Temperatures

BY G. L. PICKARD* AND F. E. SIMON
Clarendon Laboratory, Oxford
* Now at University of British Columbia, Vancouver

MS. received 5 *March* 1948

ABSTRACT. The specific heats of palladium and sodium were measured between helium and hydrogen temperatures, the specific heats of mercury between helium and liquid air temperatures. Palladium shows a very high electronic term of the specific heat. Sodium exhibits anomalous behaviour in various respects. In the case of mercury a search was made for a transition at low temperatures predicted by Bridgman, but no indication was found.

The thermodynamic functions of the three metals between absolute zero and room temperature are given in tables.

§1. INTRODUCTION

THE object of this series of measurements was to extend the known data on the atomic heats of these metals for the following individual reasons. For palladium it was desired to determine the value of the electronic specific heat below 20° K. In the case of sodium an anomalous behaviour had been suspected (Simon 1926) and experiments by Kurti and Simon using the magnetic cooling method had also suggested that the atomic heat in this region was abnormally high. It was therefore considered desirable to measure the specific heat of sodium in the helium region. For mercury the possibility of a polymorphic transition manifesting itself as a discontinuity in the atomic heat/temperature curve was to be investigated.

The experimental method in each case was to use a vacuum calorimeter with electrical heating and temperature measurement. The calorimeter was enclosed within a vacuum case in a small helium "expansion" liquefier, the sample being cooled by the condensation into the calorimeter of liquid helium which was then pumped away and the calorimetric measurements made over the required temperature range.

Since the method used was of a conventional character only an outline of the experimental method is given. The cylindrical calorimeters used were of copper of 0·2 mm. thickness, tinned to avoid leaks and supported by a 2 mm. diameter thin-walled german silver tube through which the helium was condensed and pumped. The calorimeter for the palladium measurements was 1·7 cm. in diameter and 2·1 cm. long. The electrical thermometer was of silk insulated 47 S.W.G. Eureka wire of 535 Ω resistance while the heater coil of 550 Ω resistance was of the same wire. These coils were wound round the outside of the calorimeter and fixed in position by painting with bakelite varnish which was hardened

by baking for four hours. The resistance of the thermometer coil was designed to suit the Wheatstone bridge and galvanometer available and enabled the temperature to be measured to $\pm 0{\cdot}02°$ below 20° K. The 30 S.W.G. Eureka leads passed through a heat trap connected to the helium expansion chamber to reduce heat conduction to the calorimeter.

The second calorimeter was 3·0 cm. in diameter and 3·8 cm. long to accommodate the larger specimens of mercury and sodium. The heater coil of 40 S.W.G. Eureka was of 180 Ω resistance and the 47 S.W.G. Eureka thermometer coil was of 485 Ω resistance. As measurements with the mercury were also required at temperatures well above 20° K., where the Eureka thermometer becomes insensitive, it was supplemented by a platinum resistance thermometer of 40 μ bare wire wound on a thin paper base and attached to the calorimeter by bakelite varnish. This coil had a resistance of 322 Ω at 273° K., falling to 4·5 Ω at 25° K. Arrangements were made to measure the lead resistance to the platinum coil. The temperature corresponding to the platinum resistance was determined from tables published by Clusius and Vaughen (1929). The wire used in the present experiments followed closely the values for the sample Pt–23′–1915 quoted in those tables. The Eureka thermometers were calibrated in the hydrogen and helium regions by means of the vapour pressure/temperature data for these liquids, and in the intermediate region an interpolation was made on the basis of the information obtained in a previous experiment (Pickard and Simon 1948).

The samples were cooled to about 10° K. by condensing hydrogen into the calorimeter and then pumping it away. They were further cooled by condensing helium into the calorimeter and then pumping it. When all this helium had been removed, a small quantity of helium gas was re-introduced to permit thermal exchange to take place between the calorimeter and the sample. When the temperature drift of the calorimeter had become steady the calorimetric measurements were begun. The temperature was measured at 20-sec. intervals throughout this stage of the experiments, heat being supplied for periods of one or two minutes with intervals to allow the temperature drift to be measured. The experiments were carried out in such a way that the temperature difference between the calorimeter and its enclosure was always small and the thermal isolation of the calorimeter was good. It was therefore unnecessary to use any special method to correct for imperfect heat insulation. The rise in temperature due to the heating was determined by plotting a temperature/time curve before and after the heating and extrapolating to the middle of the heating period, thus correcting for the temperature drift.

During the heating the temperature of the calorimeter rose above that of its contents, but thermal equilibrium was rapidly attained after switching off the heater current. It was found that the use of different heater currents introduced no systematic error into the results. The "overheating" was of the order of $\frac{1}{2}°$ with the mercury sample; with the palladium it was less than this, and with the sodium it was so small as to be hardly noticeable.

Correction was made for the heat capacity of the calorimeter using the values for the specific heat of copper and tin measured at Leiden (Kok and Keesom 1936, Keesom and Ende 1932). The heat capacity of the resistance coils was negligible and that of the helium exchange gas could be neglected at all except the lowest temperatures.

The metals will now be dealt with individually, any special details in technique being described with reference to the particular metal concerned.

§2. PALLADIUM

Keesom and Clark (1935) have shown that the atomic heat of nickel at low temperatures contains a linear term $0{\cdot}001744\,T$ which is many times bigger than the contribution from free electrons to be expected from the Sommerfeld theory. Mott (1936) has suggested that this term is due to the unfilled quantum states or positive holes in the *d*-shell of nickel, and that a similar term should occur for the other transition metals, whether ferromagnetic or paramagnetic.

In order to test this theory and to make more quantitative data available on the subject, the specific heat of palladium was determined, palladium being chosen on account of its similarity to nickel and its high paramagnetism which, according to Mott, should imply a large electronic specific heat. Though the electronic component in these transition metals is larger than usual, it is nevertheless small

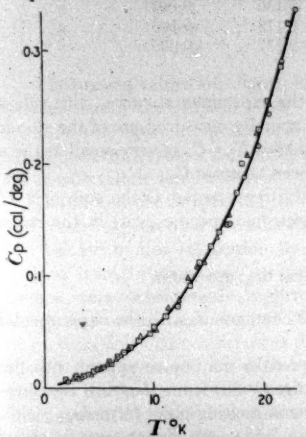


Figure 1. Atomic heat of palladium. Curve represents equation $C = 0{\cdot}0000224T^3 + 0{\cdot}0031T$. ⊙ △ ⊡ Pickard and Simon.

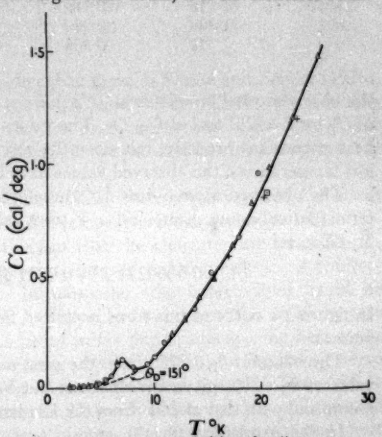


Figure 2. Atomic heat of sodium. ⊙ △ Pickard and Simon; + Simon and Zeidler.

in absolute magnitude. But since it varies in direct proportion to the temperature, whereas the specific heat due to the lattice vibrations varies near the absolute zero with the cube of the temperature, the two become comparable in magnitude at sufficiently low temperatures as pointed out by Simon (1928). Experiments were therefore carried out in the liquid hydrogen/helium region.

The sample of palladium was kindly put at our disposal by Professor N. F. Mott and was a solid cylinder of the metal, 1·4 cm. in diameter and 1·4 cm. long, weighing 25·21 gm. The palladium sample was strongly heated *in vacuo* to determine if it contained any absorbed hydrogen. The amount obtained was scarcely detectable and certainly too small to affect the results at all. Its atomic heat was measured in the small calorimeter in the range 2–22° K.

The values obtained for the atomic heat are plotted in figure 1. A mean curve was drawn and was analysed by trying a series of values of the Debye

characteristic temperature (θ_D) to calculate the lattice atomic heat (C_L) and finding which gave the most nearly linear remainder for the electronic specific heat (C_E). The best fit was found with $\theta_D=275°$, and table 1 gives values of

Table 1. Atomic heat of palladium (in cal/deg. gm. atom)

Temp. (° K.)	C_{obs}	$C_L(\rho=275°)$	C_E/T
2	0·0066	0·0002	0·0032
4	0·0132	0·0014	0·00295
6	0·023	0·0046	0·0031
8	0·036	0·0114	0·0031
10	0·054	0·0224	0·00315
12	0·076	0·039	0·0031
14	0·106	0·062	0·00315
16	0·142	0·0915	0·00315
18	0·186	0·130	0·0031
20	0·240	0·178	0·0031
22	0·306	0·237	0·00315

the observed total atomic heat (C_{obs}) from the experimental curve, with values of C_L for $\theta=275°$ and of C_E/T. The values actually measured are of the atomic heat at constant pressure, but since the correction C_p-C_v is very small at these low temperatures, the observed values have been taken as C_v.

The observed atomic heat can therefore be represented as the sum of a T^3 term (lattice atomic heat) and a T term (electronic specific heat) in the range 2°–22° K. as

$$C=0{\cdot}0000224\,T^3+0{\cdot}0031\,T \text{ cal/deg. gm. atom.}$$

In figure 1 a curve of this form is plotted for comparison with the experimental points.

The value of $\theta_D=275°$ from the present results may be compared with the value of $\theta_D=270°$ given by Meissner and Voigt (1930) from electrical measurements and with that of 260° from the Lindemann melting point formula.

In the expression for the atomic heat at 2·5° K. the linear term is about twenty times as great as the T^3 term while the two become of equal magnitude at about 12° K. The value for the electronic specific heat is, as predicted by Mott, much higher than that of nickel ($C_E=0{\cdot}001744\,T$) while work by Kok and Keesom (1936) has shown that another transition metal, platinum, also exhibits a large electronic specific heat which, however, is not as great ($C_E=0{\cdot}001607\,T$) as that for palladium.

A short note on these experiments has been published in *Nature* (Pickard 1936).

A short time ago Clusius and Schachinger (1947) measured the specific heat of palladium between 14° K. and room temperature and they found very good agreement with our experiments in the overlapping region. In addition, these authors have now all the data available to analyse the whole specific heat curve (including measurements of other authors up to 1000° K.), and in combining their values with those of the paramagnetic susceptibility they are now in a position to estimate the electronic specific heat over the whole temperature range between absolute zero and 1000°.

§3. SODIUM

The sodium used was from a sample of pure metal, hydrogen free, which had been prepared and used for experiments with neutrons in this laboratory. 18·90 gm. of sodium in the form of a cylinder which just fitted within the large calorimeter was used and measurements were made in the range 2–25° K.

The experimental points are shown graphically in figure 2 while some values for the atomic heat are given in table 2.

Table 2. Atomic heat of sodium

Temp. (° K.)	C_v	θ_D (°)	Temp. (° K.)	C_v	θ_D (°)
2	0·0055	88	10	0·140	149
3	0·0080	116	12	0·250	147
4	0·0135	130	14	0·385	148
5	0·027	129	16	0·54	150
6	0·059	120	18	0·71	152
7	0·124	109	20	0·90	154
8	0·093	137	25	1·45	156
9	0·100	150			

The present values are slightly higher than those of Simon and Zeidler (1926), who measured the specific heats between 17 and 60° K. Their values, which are shown in figure 2 together with the new ones, correspond to a θ of 159°. It can be seen from table 2 that the characteristic temperature becomes progressively lower with falling temperature.

Doubt had been expressed by Simon (1926) whether the value of θ of about 160° really represented the characteristic temperature due to lattice vibrations. He was led to this suggestion by the fact that the characteristic temperatures derived from Lindemann's melting point formula and from electrical conductivities were considerably higher. In particular, the latter values could be interpreted on the basis of a θ of 208°. Simon had suggested that, as in the case of grey tin and other substances discussed in his paper, this specific heat curve should be interpreted by a Debye term ($\theta = 202°$) plus a "Schottky" term ($\theta = 95°$) due to an internal transition of as yet unknown nature. Later, when Blackman (1934, 1935) had shown that the Debye function was less well founded than had previously been assumed, the opinion had been expressed that there was no need to assume the existence of such internal transitions, and the incompatibility of the specific heat values with the Debye curve was attributed to the defects of the latter theory.

While it is certainly possible that some relatively small anomalies in specific heat curves are due to this, we maintain the opinion that in the majority of cases discussed in the paper mentioned above, we are faced with internal transitions. A systematic investigation of the whole question has now been started at the Clarendon Laboratory, including measurements of thermal expansion, which will provide an independent value of θ. In the case of sodium, recent very careful experiments by MacDonald and Mendelssohn (1948) had shown that the electrical conductivity down to 2° K. can be represented very accurately with a θ of 202°. The difference between this value and the one derived from the specific heats without the assumption of an anomaly is much too big to be explained by a deficiency of the Debye theory.

The most striking feature in the specific heat curve is the anomaly at about 7° (see figure 3). An anomaly of a similar type had been observed by Cristescu

and Simon (1934) in beryllium at 11°. In both cases the total energy due to this anomaly is not of the order RT, but in the case of sodium $RT/70$ and in the case of beryllium $RT/700$. These are of the same order as the anomalies in the specific heat observed in superconductors, and investigation of the electrical properties (see also Kurti and Simon 1935) may shed more light on the nature of this anomaly.

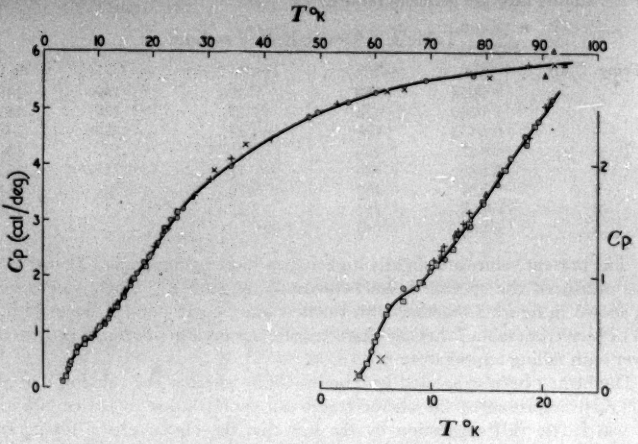


Figure 3. Atomic heat of mercury.
⊙ △ ⊡ Pickard and Simon ; + Simon ; × Pollitzer ; ⨯ Onnes and Holst.

§4. MERCURY

The object of the experiments was to investigate the possibility that a polymorphic transition, as predicted by Bridgman (1935), might occur in the region below liquid oxygen temperatures. It would manifest itself in the form of a discontinuity in the total heat/temperature curve. It was at the same time desired to extend the previous measurements (Pollitzer 1911, 1913, Simon 1923 a, 1923 b) down to the helium region.

Bridgman's measurements were concerned with the polymorphic transitions of elements at very high pressures. An extrapolation of the data obtained by him for the temperature variation of the transition pressure between the polymorphic forms of mercury indicated that this transition might occur at normal pressure at a low temperature. From an inspection of Simon's measurements (1923 a, 1923 b) he claimed that a distinct break in the tangent to the atomic heat/temperature curve was evident between 53°–68° K. and that this might be evidence of the occurrence of the transition.

The rather complicated crystal structure of mercury makes it improbable that its specific heat could be represented by a single characteristic temperature; Simon (1923 b) had found good agreement with the experimental data, using a two-term expression consisting of a Debye-function and an Einstein-function, the latter having a very low characteristic temperature, and, therefore, being responsible for the hump in the solid hydrogen region. A change in the curvature

was therefore to be expected. In view of Bridgman's prediction of a transition, however, it seemed desirable to check the specific heat again in the region between hydrogen and air temperatures.

The method adopted in the search for any thermal discontinuity was to heat the calorimeter slowly and continuously by passing a constant current through the heater coil, and to measure and record the temperature at the usual 20-sec. intervals. A graph of temperature against time was then plotted and examined for evidence of any sudden change which would have been indicated by a discontinuity in the curve.

The sample of very pure mercury (75·80 gm.) was kindly lent to us by Dr. H. London and was contained in a vessel of thin copper, varnished to prevent amalgamation of the mercury with the copper. This vessel was placed in the large calorimeter. The measurements of the atomic heat were extended over the range from 3·5° to 95° K. The slow continuous heatings extended over the whole region from 25°–80° K.

The experimental points for the atomic heat are given in figure 2 and a mean curve is drawn through them. Some values from this mean curve are given in table 3.

Table 3. Atomic heat of mercury (C_v)

Temp. (° K.)	C_v	Temp. (° K.)	C_v
3·5	0·12	25	3·18
4	$0{\cdot}20_5$	30	3·68
6	0·68	40	4·41
8	$0{\cdot}87_5$	50	4·95
10	1·06	60	5·30
12	1·31	70	5·52
15	1·72	80	5·67
20	2·46	90	5·80

The present values for the atomic heat between 10° and 14° K. and above 25° K. are slightly lower than the previous ones of Simon (1923 a, 1923 b) which are shown in figure 3 together with the results of Pollitzer (1911, 1913) and two early results of Kamerlingh Onnes and Holst (1914). The present measurements confirm that the atomic heat is anomalously high at low temperatures. The results employing continuous heating were carefully examined but *no* evidence of any thermal discontinuity was observed within the limits of the experimental accuracy, about 0·02 cal/gm. atom in the region investigated. It must be mentioned that although the temperature was only *recorded* every 20 sec. it was actually measured almost continuously, and so the possibility of even a very small discontinuity passing unnoticed is very remote.

It should be added that mercury becomes superconducting at 4·2° and that there is a small discontinuity in the specific heat at this point (Daunt and Mendelssohn 1937).

§5. THERMODYNAMIC FUNCTIONS

Our measurements at the lowest temperatures together with those of other observers at higher temperatures now permit calculation of the thermodynamic data of the three metals. Tables 4 to 6 present in the usual manner the specific heat (C_p) entropy (S), enthalpy (I) and thermodynamic potential at constant pressure (F_p) at intervals of 10°. We are very much obliged to Mr. G. K. White for the trouble he has taken in calculating these data.

Table 4. Palladium (1 gm. atom = 106·7 gm.)

T(° K.)	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$	T(° K.)	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$
0	0·0	0·0	0·0	0·0	160	5·43	5·39	491	372
10	0·054	0·038	0·110	0·27	170	5·53	5·72	545	428
20	0·24	0·121	1·40	1·02	180	5·63	6·04	601	486
30	0·66	0·30	5·71	3·27	190	5·72	6·35	658	548
40	1·30	0·57	15·3	7·54	200	5·79	6·64	715	613
50	1·93	0·93	31·2	15·15	210	5·86	6·93	774	681
60	2·56	1·34	53·6	26·5	220	5·92	7·20	833	752
70	3·10	1·77	81·9	42·1	230	5·97	7·47	892	825
80	3·56	2·22	115	62·1	240	6·02	7·72	952	901
90	3·94	2·66	153	86·1	250	6·06	7·97	1012	980
100	4·26	3·09	194	115	260	6·10	8·21	1073	1061
110	4·54	3·51	238	148	270	6·14	8·44	1134	1144
120	4·79	3·92	285	185	280	6·18	8·66	1196	1230
130	5·00	4·31	334	226	290	6·22	8·88	1258	1317
140	5·16	4·68	385	271	300	6·27	9·09	1320	1407
150	5·30	5·05	437	320					

Table 5. Sodium (1 gm. atom = 23·0 gm.)

T(° K.)	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$	T(° K.)	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$
0	0·0	0·0	0·0	0·0	160	6·00	8·40	665	679
10	0·135	0·067	0·452	0·118	170	6·07	8·76	725	764
20	0·905	0·38	5·23	2·31	180	6·13	9·11	786	854
30	2·00	0·96	19·9	8·81	190	6·18	9·44	848	947
40	2·99	1·67	45·0	21·9	200	6·24	9·76	910	1043
50	3·82	2·44	79·4	44·6	210	6·30	10·07	972	1142
60	4·36	3·19	120·6	70·7	220	6·36	10·31	1036	1233
70	4·75	3·89	166·6	105·6	230	6·41	10·60	1100	1338
80	5·02	4·54	215	148	240	6·47	10·87	1164	1445
90	5·24	5·15	267	196	250	6·53	11·14	1229	1556
100	5·42	5·71	320	251	260	6·59	11·40	1294	1668
110	5·56	6·23	375	310	270	6·65	11·65	1360	1784
120	5·66	6·72	431	375	280	6·71	11·89	1437	1892
130	5·75	7·17	488	444	290	6·77	12·13	1505	2012
140	5·84	7·60	546	518	300	6·83	12·36	1573	2134
150	5·92	8·01	605	596					

Table 6. Mercury (1 gm. atom = 200·6 gm.)

T(° K.)	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$	T(° K.)	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$
0	0·0	0·0	0·0	0·0	170	6·31	12·36	847	1254
10	1·06	0·70	4·62	2·26	180	6·36	12·72	911	1379
20	2·49	1·85	22·2	14·7	190	6·41	13·06	974	1507
30	3·68	3·11	53·6	39·5	200	6·47	13·39	1039	1640
40	4·41	4·27	94·3	76·5	210	6·53	13·71	1104	1775
50	4·95	5·32	141	125	220	6·62	14·01	1169	1914
60	5·29	6·25	193	182	230	6·74	14·31	1236	2055
70	5·51	7·08	247	249	234·4*	6·78	14·43	1265	2116
80	5·67	7·83	303	324	234·4†	7·00	16·84	1820	2116
90	5·79	8·51	360	405	240	6·95	17·01	1854	2221
100	5·89	9·12	419	494	250	6·90	17·28	1924	2418
110	5·98	9·69	478	588	260	6·85	17·56	1991	2570
120	6·05	10·21	538	687	270	6·76	17·81	2058	2746
130	6·10	10·7	599	792	280	6·70	18·06	2115	2939
140	6·16	11·15	660	901	290	6·65	18·29	2192	3107
150	6·21	11·58	722	1014	300	6·65	18·52	2258	3292
160	6·26	11·98	784	1132					

* Solid. † Liquid.

REFERENCES

BLACKMAN, M., 1934, *Proc. Roy. Soc.* A, **148**, 365, 384 ; 1935, *Ibid.*, **149**, 117, 126.
BRIDGMAN, P. W., 1935, *Phys. Rev.*, **48**, 898.
CLUSIUS, K., and SCHACHINGER, L., 1947, *Z. Naturforschung*, **2 a**, 90.
CLUSIUS, K., and VAUGHEN, J. V., 1929, *Z. ges. Kalte. Ind.*, **36**, 215.
CRISTESCU, S., and SIMON, F., 1934, *Z. phys. Chem.* B, **25**, 273.
DAUNT, J. G., and MENDELSSOHN, K., 1937, *Proc. Roy. Soc.* A, **160**, 127.
KEESOM, W. H., and CLARK, C. W., 1935, *Physica*, **2**, 513.
KEESOM, W. H., and VAN DEN ENDE, J. N., 1932, *Leiden. Comm.*, 219 b.
KOK, J. A., and KEESOM, W. H., 1936, *Leiden Comm.*, 245 a ; 1936, *Physica*, **3**, 1035.
KURTI, N., and SIMON, F., 1935, *Proc. Roy. Soc.* A, **151**, 610.
MACDONALD, D. K. C., and MENDELSSOHN, K., 1948, *Nature, Lond.*, in press.
MEISSNER, W., and VOIGT, B., 1930, *Ann. Phys., Lpz.*, **7**, 893.
MOTT, N. F., 1936, *Proc. Roy. Soc.* A, **152**, 42.
ONNES, H. KAMERLINGH, and HOLST, G., 1914, *Leiden Comm.*, 142 c.
PICKARD, G. L., 1936, *Nature, Lond.*, **138**, 123.
PICKARD, G. L., and SIMON, F. E., 1948, *Proc. Phys. Soc.*, **60**, 405.
POLLITZER, F., 1911, *Z. Elektrochem.*, **17**, 5 ; 1913, *Ibid.*, **19**, 513.
SIMON, F., 1923 a, *Ann. Phys., Lpz.*, **68**, 241 ; 1923 b, *Z. phys. Chem.* A, **107**, 279 ; 1926, *Ber. Pr. Akad. Wiss.*, **33**, 477 ; 1928, *Z. Elektrochem.*, **34**, 530.
SIMON, F., and ZEIDLER, W., 1926, *Z. phys. Chem.* A, **123**, 383.

The Theory of the Propagation of Second Sound in Helium II

BY R. B. DINGLE,
H. H. Wills Physical Laboratory, Royal Fort, Bristol

Communicated by N. F. Mott ; MS. received 20 *January* 1948

ABSTRACT. In this paper the conduction of heat in liquid helium II will be discussed, and the velocity of second sound, the attenuation, and reflection coefficients under various conditions will be obtained. The shape of the velocity–temperature curve will be discussed on the basis of Landau's theory of helium II, and Tisza's theory will be criticized.

§1. INTRODUCTION

THE experiments of Peshkov (1946) in Russia and Lane and co-workers (1947) in America have demonstrated beyond doubt the existence of a new type of sound wave in liquid helium II, one in which temperature fluctuations are the important phenomena and not pressure fluctuations as in the case of ordinary sound. The essential assumptions of the theories developed by Tisza (1938, 1940 and 1947) and Landau (1941, 1944) which account for second sound are as follows: (i) Helium II may, at least from a phenomenological standpoint, be divided into two parts. One of these parts, the so-called superfluid, is probably to be explained as some kind of condensed phase, perhaps the condensed phase to be expected below a certain temperature in a substance obeying Bose-Einstein statistics (London 1938, 1939). (ii) The entropy of one part, the so-called normal fluid, is greater than that of the superfluid. Kapitza (1941) has shown, by experiments in which the two fluids are separated by the use of a very narrow gap, that the entropy of the superfluid is very small: it may be zero. (iii) To these assumptions may be added a third, which is important in the theory of the flow of helium II. The normal fluid obeys the hydrodynamical equations appropriate to any other normal

fluid; thus its equations contain a coefficient of viscosity; but the superfluid obeys hydrodynamical equations appropriate to a fluid in which the curl of the velocity is everywhere zero, so that the motion of the superfluid is always irrotational. In particular, the tangential velocity of the normal fluid must vanish at a surface, whilst the tangential velocity of the superfluid need not vanish there.

In this paper we shall consider several problems in the theory of the propagation of heat and of second sound in helium II using the above assumptions; the problems are chosen both because of their importance in experimental investigations and also as illustrations of the use of the equations and boundary conditions which will be established in this paper.

§ 2. THE SHAPE AND SPEED OF ORDINARY "CONDUCTED" PULSES

Throughout this section, the word "conducted" will mean a flow of heat not involving fluid movements.

When Tisza (1938, 1940) predicted the existence of "temperature waves", he pointed out that these waves could easily be distinguished from ordinary "conduction waves" by the fact that the latter would be greatly attenuated, and would also have a velocity which would be proportional to the square root of the frequency, unlike the temperature waves, which should have a velocity independent of the frequency.

For suppose that T is the temperature excess over the initial uniform value, t the time, K the ordinary thermal conductivity (due to heat flow without fluid movements), and C the specific heat at constant volume. Then elementary arguments give the equation of heat conduction as $dT/dt=(K/\rho C)\nabla^2 T$, where ρ is the density. Suppose for simplicity that we are dealing with plane propagation along the z axis. Then this equation becomes $dT/dt=Dd^2T/dz^2$ where $D=K/\rho C$. If the source consists of a plane at $z=0$ undergoing sinusoidal temperature fluctuations of frequency $\omega/2\pi$, then $dT/dt=i\omega T$; $d^2T/dz^2=i\omega T/D$; and $T\propto e^{i\omega t}\exp\{-z(1+i)\sqrt{(\omega/2D)}\}$.

Hence the distance in which the amplitude of temperature falls to $1/e$ of the temperature of the source is just one wavelength, and the apparent velocity is $\sqrt{(2D\omega)}$.

If on the other hand the "conduction wave" is emitted from the plane $z=0$ at the time $t=0$ as a sharp pulse, the appropriate solution of the equation is

$$T\propto\exp(-z^2/4tD)/\sqrt{t},$$

where $D=K/\rho C$. If the conduction effect is the only one of importance, or if all the temperature waves of the Tisza-Landau type are reflected back without loss, then the proportionality constant may be found by integration of T (the excess temperature) over all space. The result must clearly be independent of time except during the initial sharp emission of the pulse. Thus if Q is the quantity of heat in the pulse,

$$Q=C\int_{-\infty}^{\infty}T\,dz=\frac{AC}{\sqrt{t}}\int_{-\infty}^{\infty}\exp(-z^2/4tD)\,dz=AC\sqrt{(4\pi D)},$$

where A is a proportionality constant. Thus finally

$$T=\frac{Q}{C\sqrt{(4\pi tD)}}\exp(-z^2/4tD).$$

The maximum of the pulse, for a given value of x, is given by $(\partial T/\partial t)_x = 0$; i.e. $t_{max} = x^2/2D$ and $T_{max} = 0{\cdot}242Q/Cx$. Thus the apparent velocity of the pulse depends on the distance x; this is simply because the conduction effect is dispersive. The shape of this "conducted pulse" may be deduced from the above equations.

In problems of this type it is only necessary to obtain solutions of the physical problem involved for the simple case of pure sine waves; all other cases may be deduced mathematically from this solution by means of Fourier analysis.

§3. THE VELOCITY OF SECOND SOUND

A satisfactory deduction of the velocity of second sound was first given by Landau (1941). Recently a simplified approximate treatment has been proposed by Gogate and Pathak (1947). They obtain the same formula for the velocity, but some of their arguments appear unsatisfactory to the present author, especially as in the deduction of one equation second-order quantities (involving the product of the two first-order quantities dT and v_n in their notation) are used. An alternative simplified theory is therefore given below, and a more detailed discussion of the problem in §9. We consider that the treatment given by Meyer and Band (1947) is illogical, as they commence by writing down equations appropriate to a single fluid, and finally divide it into two effective parts.

We may form the following physical picture of the wave motion associated with second sound: due to the relative motion of normal fluid which carries entropy, and superfluid which does not, we obtain a state of the liquid such that alternate layers contain an excess or a deficit of entropy. It follows that these layers are at an effectively higher and lower temperature respectively. In the following simplified deduction of the velocity of such waves, we shall neglect the correction caused by the small pressure differences which will be set up in the liquid due to the fact that the layers at the higher temperatures will contract slightly, and the layers at the lower temperature will expand a little, helium II having a negative coefficient of thermal expansion. We shall also neglect the difference between total and partial differentiation with respect to time, since this difference is quadratic in the velocities—that is, we neglect the difference between differentiation with respect to time whilst remaining stationary in space, and differentiation whilst following the movements of the particles.

Let T be the temperature excess over the mean undisturbed value, ρ the density, S the entropy per unit mass of the liquid considered as a whole, ρ_n the density of normal fluid, ρ_s the density of superfluid, t the time, V_n the velocity of the normal fluid, V_s the velocity of the superfluid, and ϕ the thermodynamic potential of the fluid per unit mass.

Then $\rho = \rho_n + \rho_s$, and for the current j we have $j = \rho_n V_n + \rho_s V_s$. If we assume for the present that entropy is carried only by the normal fluid (we return to this question in §9), we have the equation of conservation of entropy,

$$\partial(\rho S)/\partial t + S\nabla.(\rho V_n) = 0.$$

Since $C = T'\partial S/\partial T$, this equation may be rewritten (provided the fluctuations are small and density changes negligible)

$$\partial T/\partial t + (T'S/C)\nabla.V_n = 0. \qquad \ldots\ldots(1)$$

Here T' is the actual temperature of the fluid.

Confining attention to waves in which the fluctuations in pressure and density are negligible (these pressure fluctuations being associated with ordinary first

sound), we see from the conservation equation $\partial\rho/\partial t+\nabla.j=0$ that $j=0$, as any other constant would imply a constant net flow of the fluid in one direction.

The final equation required is provided by an argument similar to that given by Landau (1941), deducing the equation giving the relation between force and acceleration for the superfluid alone. We imagine two thermally isolated vessels, connected by a capillary so narrow that only superfluid may pass along it. Then the net work required to transfer unit mass of superfluid from one vessel to the other will equal the difference $\partial\phi$ in the thermodynamic potential (Gibbs function) per unit mass between the two vessels. This work will give the change in potential energy. Thus if we consider again the bulk liquid, the gradient of the thermodynamical potential with respect to distance (taken with a negative sign) will give the effective force acting on the superfluid alone. Therefore, for the superfluid, the equation relating acceleration to force is given by $\partial V_s/\partial t+\nabla\phi=0$. Since $\nabla\phi=-S\nabla T+V\nabla p=-S\nabla T$ for displacements where ∇p is negligible, $\partial V_s/\partial t-S\nabla T=0$. Equating j to zero, we have $\rho_n V_n=-\rho_s V_s$, and hence

$$\partial V_n/\partial t+S\rho_s\nabla T/\rho_n=0. \qquad \ldots\ldots(2)$$

Combining this with equation (1), we obtain a wave equation

$$\frac{\partial^2 T}{\partial t^2}=\frac{T'S^2}{C}\frac{\rho_s}{\rho_n}\nabla^2 T, \qquad \ldots\ldots(3)$$

with which there is associated a velocity $u_2=(T'S^2\rho_s/C\rho_n)^{\frac{1}{2}}$.

We shall now show that in a travelling plane wave of second sound the heat flow at any point is proportional to the temperature at that point, and not to the temperature gradient as in a conducted wave, and also that in a stationary plane wave of second sound the heat flow at any point and at any time is $\frac{1}{2}\pi$ out of phase both in time and in space with the temperature at that point. These results follow from equation (2). The rate of heat flow per unit area is $T'\rho SV_n$, and thus the rate of heat flow is proportional to the velocity of the normal fluid. In a travelling wave we may take T proportional to $\cos(\omega t-\gamma z)$, where γ is the wave number ω/u_2. Thus $\partial V_n/\partial t\propto\partial T/\partial z\propto\sin(\omega t-\gamma z)$, and hence $V_n\propto\cos(\omega t-\gamma z)\propto T$. In a stationary wave we may assume T proportional to $\cos\omega t\cos\gamma z$ with suitable origins of time and space. V_n is then proportional to $\sin\omega t\sin\gamma z$. Thus the velocity of the normal fluid, to which the rate of heat flow is proportional, is $\frac{1}{2}\pi$ out of phase with the temperature both in space and in time.

§4. THE ENERGY FLOW IN SOUND

Energy considerations are valuable in the deduction of reflection coefficients and allied problems, and it is therefore important to determine the rate of energy flow per unit area in sound waves.

In a travelling wave of first sound, the kinetic energy per unit volume is given by $\frac{1}{2}\rho V^2$, and if the wave is sinusoidal, the average density of kinetic energy is given by $\frac{1}{4}\rho(V^0)^2$, where V^0 is the peak velocity. It is easily verified that the density of potential energy is the same as that of kinetic energy in a plane wave of first sound (§10); it follows that the total energy density is $\frac{1}{2}\rho(V^0)^2$. Therefore the rate of energy flow of first sound per unit area is given by $\frac{1}{2}\rho(V^0)^2u_1$, where u_1 is the velocity of the waves. Since in first sound the maximum excess pressure p^0 is given in terms of the maximum velocity V^0 by the equation $p^0=\rho u_1 V^0$, we may also write for the rate of energy flow $\frac{1}{2}(p^0)^2/\rho u_1$.

In a travelling wave of second sound, the kinetic energy per unit volume is $\frac{1}{2}(\rho_n V_n^2+\rho_s V_s^2)$. Since $\rho_n V_n=-\rho_s V_s$ this may be rewritten as $\frac{1}{2}\rho_n\rho V_n^2/\rho_s$ or $\frac{1}{2}\rho_n\rho(V_n^0)^2/\rho_s$, where V_n^0 is the peak value of the velocity of the normal fluid. It will be proved later (§ 10) that the density of potential energy has the same value, and hence we may deduce that the total energy density is given by $\frac{1}{2}\rho_n\rho(V_n^0)^2/\rho_s$. Thus the rate of energy flow of second sound per unit area is $\frac{1}{2}\rho_n\rho(V_n^0)^2u_2/\rho_s$, where u_2 is the velocity of this sound. Since in second sound the maximum excess temperature is given by $T^0=\rho_n u_2 V_n^0/\rho_s S$, we may also write for the rate of energy flow $\rho\rho_s S^2(T^0)^2/2\rho_n u_2=\rho u_2 C(T^0)^2/2T'$, where T' is the mean temperature of the liquid.

§ 5. THE ATTENUATION OF SECOND SOUND DUE TO OTHER HEAT PROCESSES

It is interesting to attempt to calculate the attenuation of second sound caused by non-reversible conduction of heat. For the sake of simplicity only processes in which heat flow is directly proportional to the temperature gradient (with a factor of proportionality K per unit cross-section per unit time) will be considered. Thus included in K will be, for example, the ordinary heat conductivity without fluid movement.

In the equation of entropy conservation we must now consider also irreversible processes. If we consider unit volume, the rate of loss of entropy is $-\partial(\rho S)/\partial t$. This must be equated to the sum of the reversible loss $\rho S\nabla.V_n$ (assuming that only the normal phase carries entropy), and the irreversible loss $(1/T')(\partial Q/\partial t)=-(K/T')\nabla^2 T$, where K is the irreversible conductivity defined above. Thus $\partial(\rho S)/\partial t+\rho S\nabla.V_n-(K/T')\nabla^2 T=0$ and, therefore,

$$\frac{\partial T}{\partial t}+\frac{T'S}{C}\nabla.V_n-\frac{K}{\rho C}\nabla^2 T=0.$$

Eliminating V_n between this and equation (2), which is unchanged to this approximation, we find for the (complex) velocity u_2,

$$u_2^2=\frac{T'S^2\rho_s}{C\rho_n}\left(1+\frac{Ki\omega\rho_n}{\rho T'S^2\rho_s}\right).$$

Now the amplitude of the wave is proportional to $\exp(i\omega z/u_2)$, of which the real part decays as $\exp(-K\omega^2 z/2u_2^3\rho C)$. Thus the distance in which the amplitude falls to $1/e$ of its initial value is given by $2u_2^3\rho C/K\omega^2$. Since $u_2=0$ at the λ-point, we see that the attenuation increases rapidly in the region of the λ-point. It also increases near the absolute zero since here C becomes very small, whilst u_2 should remain finite according to theory (§ 11). Thus the attenuation should exhibit a minimum between 0° K. and the λ-point.

For the ordinary non-reversible thermal conductivity K it is reasonable to insert the value found for helium I, where it is not masked by second-sound convection effects. Keesom and Keesom (1936) give $K\sim 6\times10^{-5}$ cal/deg.cm.sec. at 3·3° K. For the specific heat we insert the smoothed values given by Keesom (1942). Then very approximately the distance in which the amplitude falls to $1/e$ of its initial value is $2\times10^{30}/f^2$ cm. at 1·6° K. and $7\times10^{10}/f^2$ cm. at 2·18° K., falling rapidly to zero at 2·19° K., the λ-point. It therefore appears likely that the attenuation due to irreversible heat effects is only of importance at very high frequencies except close to the λ-point and probably, also, at extremely low temperatures. In the above, f is the frequency in cycles/sec.

§6. THE REFLECTION OF SECOND SOUND AT WALLS HAVING FINITE HEAT CONDUCTIVITY

The solution of this problem of reflection is important because one experimental method for measuring the velocity of second sound depends on the setting up of standing waves in a tube. Suppose that there is a source of temperature waves at one end of the tube, and a solid block of material (e.g. a thermometer) of thermal conductivity K at the other. As the heat conductivity, or rather heat convection, of helium II is so much greater than that of any solid, we shall assume that the solid block is effectively infinite in depth. In analogy with transmission-line problems we consider the problem of reflection at a junction of two "transmission tubes", the liquid being one and the solid the other. Let the temperature T be the counterpart of voltage, and the heat flux I the counterpart of current. The ratio T/I across a junction will be called the impedance Z across that junction. Then the problem may be divided into three parts: first, the calculation of the characteristic impedance Z_0 of the liquid; second, the calculation of the impedance Z_l of the solid at the interface between solid and liquid; and finally, the deduction of the reflection and transmission coefficients from these impedances.

For simple harmonic waves of frequency $\omega/2\pi$, equations (1) and (2) of §3 may be written $\partial I/\partial x = -i\omega C\rho T$ and $\partial T/\partial x = -i\omega I/\rho C u_2^2$, where I is the heat flux, $\rho S T' V_n$.

Now for a loss-free transmission line of inductance L^* per unit length and capacity C^* per unit length, we have as equations linking voltage V and current I

$$\partial I/\partial x = -i\omega C^* V \qquad \ldots\ldots(4)$$

and

$$\partial V/\partial x = -i\omega L^* I. \qquad \ldots\ldots(5)$$

Comparing these equations with those for our case, we may put $C^* = C\rho$ and $L^* = 1/\rho C u_2^2$. (From §5 we may also derive the value $G = \omega^2 K_{liq}/u_2^2$ for the effective resistance per unit length *across* the tube.) We may now take over the well known theory of loss-free transmission lines. For instance, we may write the characteristic impedance of the liquid as $Z_0 = \sqrt{(L^*/C^*)} = 1/\rho C u_2$. The temperature $T(x)$ at a point x and the heat flux $I(x)$ are given by

$$T(x) = T_i e^{-i\beta x} + T_r e^{i\beta x}, \qquad \ldots\ldots(6)$$

$$Z_0 I(x) = T_i e^{-i\beta x} - T_r e^{i\beta x}, \qquad \ldots\ldots(7)$$

where T_i and T_r are constants giving the temperature maxima of the forward and reflected waves respectively, and β is the wave number ω/u_2. These formulae satisfy equations (4) and (5).

The standing wave ratio (i.e. the ratio of maximum to minimum) will be given by $(|T_i| + |T_r|)/(|T_i| - |T_r|)$. Using (6) and (7) to determine this in terms of Z_0 and Z_l (the impedance at the end $x = l$), we have for the standing wave ratio $|Z_l/Z_0|$. By a double application of (6) and (7) we may also show that the ratio of temperature to heat flux at a distance l from a junction, across which there is an impedance Z_l, is given by

$$\frac{T}{I} = Z_0 \left\{ \frac{Z_l \cos\beta l + iZ_0 \sin\beta l}{iZ_l \sin\beta l + Z_0 \cos\beta l} \right\}.$$

Next we calculate Z_l, the impedance of the solid at the interface. It follows from §2 that the temperature in the solid at a depth z from the plane of contact is given by

$$T_{sol} = T_0 e^{i\omega t} \exp\{-z\gamma(1+i)\},$$

where $\gamma^2 = \omega\rho_{sol}C_{sol}/2K$, ρ_{sol} and C_{sol} being respectively the density and the specific heat of the solid. The rate of flow of heat into the block per unit area is then given by $I_{sol} = -K(\partial T_{sol}/\partial z)_{z=0} = \alpha(1+i)T_0 e^{i\omega t}$, where $\alpha = K\gamma = (\frac{1}{2}\omega\rho_{sol}C_{sol}K)^{\frac{1}{2}}$. Thus the impedance at the interface is given by $Z_l = 1/\alpha(1+i)$. The standing wave ratio is $|Z_l/Z_0| = u_2\rho C/\alpha\sqrt{2}$.

Finally, we must find the reflection and transmission coefficients in terms of Z_0 and Z_l. These may be obtained with the aid of the two boundary conditions that the temperature must be continuous across the interface, and that the heat flow must be a continuous function of distance.

Taking the origin of coordinates at the interface, the first boundary condition gives $T_i + T_r = T_{sol}$, and the second $I_i + I_r = I_{sol}$. Using the relations $T_i = Z_0 I_i$, $T_r = -Z_0 I_r$, and $T_{sol} = Z_l I_{sol}$, we obtain a reflection coefficient

$$I_r/I_i = (Z_l - Z_0)/(Z_l + Z_0),$$

and a transmission coefficient

$$I_{sol}/I_i = 2Z_0/(Z_l + Z_0) \simeq 2Z_0/Z_l.$$

The (complex) transmission coefficient is thus given by $2\alpha(1+i)/\rho C u_2$.

§7. THE REFLECTION OF SECOND SOUND AT A LIQUID–VAPOUR INTERFACE

In an experimental arrangement originally suggested by Onsager, and used in the experiments of Lane and co-workers (1947), stationary plane waves of second sound are set up in a tube of liquid helium II and, owing to imperfect reflection at the liquid–vapour interface at one end, first sound is transmitted in the vapour. The process may be pictured very simply as follows: the stationary waves of second sound in the liquid cause a periodic fluctuation in temperature at the surface of the liquid, since, as there is very little heat flow at the interface, this is a temperature antinode. There is, therefore, a periodic fluctuation in the amount of vapour emitted from the surface; this produces first sound in the vapour. In this section we demonstrate how the reflection coefficient for such a process may be calculated.

Suppose that a travelling wave of second sound of unit amplitude is incident on the liquid–vapour interface, and that a wave of amplitude R is reflected, the remainder being transmitted into the vapour as first sound. The boundary conditions which will be assumed are: (i) the vapour pressure at the interface is the equilibrium pressure at the temperature of the surface of the liquid; (ii) the difference in the energies transmitted by the incident and reflected waves in the liquid is equal to the energy transmitted by the sound wave in the vapour.

In the liquid the temperature at any point is given by $T_{liq} = T_0 e^{i\omega t}(e^{-ikx} + Re^{ikx})$. Here $\omega/2\pi$ is the frequency of the wave, k the wave number ω/u_2, and R is the reflection coefficient.

At the interface we have $T_{liq} = T_0(1+R)$. Now from the first boundary condition we may deduce that the excess pressure in the first sound wave at the interface is given by $p_0 = T_0(1+R)\,\partial p/\partial T$, where $\partial p/\partial T$ is the slope of the equilibrium vapour pressure with respect to temperature; this is given by the Clausius-Clapeyron equation $\partial p/\partial T = L/T(V_{vap} - V_{liq})$.

The excess pressure at any point in the vapour is given by $p = p_0 \exp(i\omega t - ikx)$. Using the results of § 4 we deduce that the rate of energy flow per unit area in the

vapour is $\frac{1}{2}p_0^2/\rho_v u_1$, and the difference in rates between the energy flow in the incident and reflected waves in the liquid is $\frac{1}{2}\rho_l u_2 C T_0^2(1-R^2)/T'$. From equation (5)

$$\left(\frac{\partial p}{\partial T}\right)^2 \frac{(1+R)^2}{\rho_v u_1} = \frac{\rho_l u_2 C}{T'}(1-R^2),$$

whence
$$R = \left(1 - \left(\frac{\partial p}{\partial T}\right)^2 \frac{T'}{\rho_l \rho_v u_1 u_2 C}\right) \Big/ \left(1 + \left(\frac{\partial p}{\partial T}\right)^2 \frac{T'}{\rho_l \rho_v u_1 u_2 C}\right).$$

§8. THE PROPAGATION OF SECOND SOUND IN THIN TUBES

For convenience in experimental technique, the investigation of second sound is carried out in tubes of diameter about 1·5 cm. It is therefore important to enquire whether the velocity is appreciably affected by the narrowness of the tube. The physical reason of the expected effect is that the normal fluid has a viscosity and, therefore, sticks to the walls of the tube, whilst the superfluid has no viscosity and, therefore, does not stick.

Consider a plane wave propagated down a circular tube of radius r with viscosity of the normal part of the fluid μ. Then the force acting on the superfluid is that due to the gradient of the thermodynamic potential per unit mass, assuming the entropy to be zero, and the force acting on the normal fluid is that due to the sticking effect. The mass-acceleration of the fluid as a whole will be equal to the sum of the sticking force and the gradient of pressure.

By an argument similar to that given by Rayleigh (1926), we obtain the equation

$$\frac{\partial^2 \rho}{\partial t^2}\left(1 + \frac{2}{r}\sqrt{\frac{\mu}{2\omega\rho}}\right) + \frac{2}{r}\frac{\partial \rho}{\partial t}\sqrt{\frac{\mu\omega}{\rho}} = \nabla^2 p.$$

This equation should hold provided the tube is sufficiently wide for the thickness adhering to the walls to be small compared with the diameter of the tube. Provided the correction terms we are calculating are small, we may rewrite this equation as $\partial^2\rho/\partial t^2 = (1/f)\nabla^2 p$, where

$$f = 1 + (2/r)\sqrt{(\mu/2\omega\rho)} + (2/i\omega r)\sqrt{(\mu\omega/\rho)}.$$

This equation, with $\partial^2 S/\partial t^2 = S^2\rho_s\nabla^2 T/\rho_n$†, gives the velocity of the two sounds. By a slight extension of the work of Landau (1941), we obtain

$$u^4 - u^2\left\{\frac{1}{f}\left(\frac{\partial p}{\partial \rho}\right)_s + \frac{T'S^2\rho_s}{C\rho_n}\right\} + \frac{T'S^2\rho_s}{C\rho_n f}\frac{\partial p}{\partial \rho} = 0.$$

According to this equation, the velocity of first sound is given by

$$u_1 = \sqrt{\left(\frac{1}{f}\frac{\partial p}{\partial \rho}\right)} \simeq \frac{\partial p}{\partial \rho}\left(1 - \frac{1}{r}\sqrt{\frac{\mu}{2\omega\rho}}\right),$$

and the velocity of second sound by $u_2 = (T'S^2\rho_s/C\rho_n)^{\frac{1}{2}}$. Thus the correction to the velocity of second sound, on account of the narrowness of the tube, is much smaller than that to the velocity of first sound.‡

† This equation (cf. (3) § 3) is true even in the case where pressure differences are not negligible (Landau 1941).

‡ *Note added in proof.* A more detailed consideration now shows that there is a very small correction proportional to the product of the coefficient of thermal expansion and the square root of the viscosity.

§9. THE USE OF THE LAGRANGIAN AND THE TOTAL ENERGY IN THE DEDUCTION OF THE VELOCITY OF SECOND SOUND IN A FLUID CONSISTING OF PARTS WITH TWO DIFFERENT ENTROPIES

There is some question as to whether the entropy of the superfluid in helium II is really exactly zero; we shall therefore discuss and apply a method which is an extension of that of Tisza (1947).

Let the entropy of the normal fluid alone be S_n, and the entropy of the superfluid is S_s. Then the measured entropy S of the whole fluid is given by $\rho S=\rho_n S_n+\rho_s S_s$. Now we introduce new coordinates ξ and η such that $\rho\xi=\rho_n V_n+\rho_s V_s$ and $\eta=(\rho_n V_n S_n+\rho_s V_s S_s)/(\rho_n S_n+\rho_s S_s)-\xi$. Since η is defined as the excess over ξ, the velocity of the fluid as a whole, of the velocities of the normal and the superfluid, weighted according to the amount of entropy which they carry, the equation of the conservation of entropy is expressible in these new coordinates in the simple form $\partial S/\partial t+S\nabla.\eta=0$. Also in these coordinates, the equation of the conservation of mass becomes $(1/\rho)(\partial\rho/\partial t)+\nabla.\xi=0$. After some reduction, using the fact that $\rho=\rho_n+\rho_s$, we obtain for η, $S^2\rho^2\eta=\rho_n\rho_s S^*(V_n-V_s)$, where $S^*=S_n-S_s$. Thus the transfer of entropy depends on the relative velocities of the parts with different entropies, as is clear physically. From these equations we may obtain V_n and V_s in terms of ξ and η as follows:

$$V_n=\xi+(\eta\rho S/\rho_n S^*); \qquad V_s=\xi-(\eta\rho S/\rho_s S^*).$$

Hence the kinetic energy per unit volume is given by

$$\tfrac{1}{2}\rho_n V_n^2+\tfrac{1}{2}\rho_s V_s^2=\tfrac{1}{2}\rho\xi^2+\tfrac{1}{2}(\rho^3 S^2\eta^2/\rho_n\rho_s S^{*2}).$$

The potential energy is due to fluctuations both in volume and in entropy. Now if δU represents a small change in internal energy, we may expand in terms of δV, the small change in volume, and δS, the small change in entropy. Retaining terms up to the second order, we have

$$\begin{aligned}\delta U&=\left(\frac{\partial U}{\partial V}\right)_S\delta V+\left(\frac{\partial U}{\partial S}\right)_V\delta S+\tfrac{1}{2}\frac{\partial^2 U}{\partial V^2}\delta V^2+\frac{\partial^2 U}{\partial V\partial S}\delta V.\delta S+\tfrac{1}{2}\frac{\partial^2 U}{\partial S^2}\delta S^2\\&=-p\delta V+T\delta S-\tfrac{1}{2}\left(\frac{\partial p}{\partial V}\right)_S\delta V^2-\left(\frac{\partial p}{\partial S}\right)_V\delta V.\delta S+\tfrac{1}{2}\left(\frac{\partial T}{\partial S}\right)_V\delta S^2.\end{aligned}$$

Thus the minimum work to cause the fluctuations is given by (Landau and Lifshitz 1938)

$$\begin{aligned}\delta U+p\delta V-T\delta S&=-\tfrac{1}{2}\left(\frac{\partial p}{\partial V}\right)_S\delta V^2-\left(\frac{\partial p}{\partial S}\right)_V\delta V.\delta S+\tfrac{1}{2}\left(\frac{\partial T}{\partial S}\right)_V\delta S^2\\&=-\tfrac{1}{2}\left(\frac{\partial p}{\partial V}\right)_S\delta V^2-\left(\frac{\partial p}{\partial T}\right)_V\left(\frac{\partial T}{\partial S}\right)_V\delta V.\delta S+\frac{T'}{2C}\delta S^2\\&=\tfrac{1}{2}\left(\frac{\partial p}{\partial\rho}\right)_S\left(\frac{\delta\rho}{\rho}\right)^2-\rho\alpha\left(\frac{\partial p}{\partial\rho}\right)_T\frac{T'}{C}\delta V.\delta S+\frac{T'S^2}{2C}\left(\frac{\delta S}{S}\right)^2.\end{aligned}$$

Here α is the coefficient of thermal expansion, which is small, and may be neglected; the inclusion of this term gives the coupling between first and second sound. Thus the potential energy is, per unit mass,

$$\tfrac{1}{2}\left(\frac{\partial\rho}{\partial p}\right)_S\left(\frac{\delta\rho}{\rho}\right)^2+\frac{T'S^2}{2C}\left(\frac{\delta S}{S}\right)^2.$$

It is convenient to rewrite the equations of continuity and of entropy conservation in terms of displacements rather than velocities. If we suppose that ξ_1 is the displacement corresponding to the velocity ξ, and η_1 is the displacement corresponding to the velocity η, we may write the equations of continuity and of entropy conservation as

$$\delta\rho/\rho = -\nabla.\xi_1 \qquad \text{and} \qquad \delta S/S = -\nabla.\eta_1.$$

The potential energy per unit volume becomes in these coordinates

$$\frac{\rho}{2}\left(\frac{\partial p}{\partial \rho}\right)_S(\nabla.\xi_1)^2 + \frac{\rho T' S^2}{2C}(\nabla.\eta_1)^2,$$

and the kinetic energy per unit volume becomes

$$\tfrac{1}{2}\rho\dot{\xi}_1^2 + \tfrac{1}{2}\frac{\rho^3}{\rho_n\rho_s}\left(\frac{S}{S^*}\right)^2\dot{\eta}_1^2,$$

where dots denote differentiation with respect to time. Thus the total energy density is

$$\tfrac{1}{2}\rho\dot{\xi}_1^2 + \tfrac{1}{2}\frac{\rho^3}{\rho_n\rho_s}\left(\frac{S}{S^*}\right)^2\dot{\eta}_1^2 + \tfrac{1}{2}\left(\frac{\partial p}{\partial \rho}\right)_s(\nabla.\xi_1)^2 + \frac{\rho T' S^2}{2C}(\nabla.\eta_1)^2,$$

and the Lagrangian density is

$$\tfrac{1}{2}\rho\dot{\xi}_1^2 + \tfrac{1}{2}\frac{\rho^3}{\rho_n\rho_s}\left(\frac{S}{S^*}\right)^2\dot{\eta}_1^2 - \tfrac{1}{2}\left(\frac{\partial p}{\partial \rho}\right)(\nabla.\xi_1)^2 - \frac{\rho T' S^2}{2C}(\nabla.\eta_1)^2.$$

The appropriate Lagrangian equations of motion are

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\xi}_1}\right) - \frac{\partial L}{\partial \xi_1} = -\nabla\frac{\partial L}{\partial(\nabla.\xi_1)} \qquad \text{and} \qquad \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\eta}_1}\right) - \frac{\partial L}{\partial \eta_1} = -\nabla\frac{\partial L}{\partial(\nabla.\eta_1)}.$$

Here $\partial L/\partial(\nabla.\xi_1)$ and $\partial L/\partial(\nabla.\eta_1)$ may be interpreted as pressures, and the gradients of these pressures give rise to forces. We obtain then the wave-equations $\ddot{\xi} = (\partial p/\partial\rho)_S\nabla^2\xi_1$ and $(\rho^2/\rho_n\rho_s)(S/S^*)^2\ddot{\eta}_1 = (T'S^2/C)\nabla^2\eta_1$.

The velocity of the first sound associated with the first of these equations is, therefore, $u_1 = (\partial p/\partial\rho)^{\frac{1}{2}}$, and the velocity of second sound as given by the second is $u_2 = (T'\rho_n\rho_s S^{*2}/C\rho^2)^{\frac{1}{2}}$.

We must again emphasize at this point that $S^* = S_n - S_s$ is the difference in entropies between the normal and the superfluid when account is taken of the fact that the fraction of the normal fluid is ρ_n/ρ and the fraction of superfluid is ρ_s/ρ. If we could isolate the superfluid, and it had the same density ρ as the bulk fluid, we could measure its entropy and call it, say, S_s'. Similarly, if we had only normal fluid of density ρ, its entropy might be called S_n'. We should then have $\rho_s S_s = \rho S_s'$ and $\rho_n S_n = \rho S_n'$. Then

$$S^* = S_n - S_s = \rho(S_n'/\rho_n - S_s'/\rho_s).$$

In this notation

$$u_2^2 = (T'/C)\{\rho_s(S_n')^2/\rho_n + \rho_n(S_s')^2/\rho_s - 2S_n'S_s'\}.$$

This formula reduces to the well known one $u_2^2 = T'S^2\rho_s/C\rho_n$ when any entropy associated with the superfluid is ignored.

§ 10. THE AVERAGE EQUALITY OF POTENTIAL AND KINETIC ENERGY IN A PLANE WAVE OF SECOND SOUND

We may now prove the proposition of § 4 that the time averages of the potential and kinetic energies are equal. This proposition may be stated in the alternative form of the vanishing of the time-average of the Lagrangian.

The Lagrangian density is, as shown in the last section,

$$\left[\tfrac{1}{2}\rho\dot{\xi}_1^2 - \tfrac{1}{2}\rho\left(\frac{\partial p}{\partial \rho}\right)_S(\nabla.\xi_1)^2\right] + \left[\tfrac{1}{2}\frac{\rho^2}{\rho_n\rho_s}\left(\frac{S}{S^*}\right)^2\dot{\eta}^2 - \tfrac{1}{2}\frac{T'S^2\rho}{C}(\nabla.\eta_1)^2\right],$$

the terms in the first bracket being contributed by first sound and those in the second by second sound. If we suppose that ξ_1 and η_1 are displacements in plane sine waves, we shall have

$$\xi_1 = \xi_1^0 \sin\omega(t - x/u_1) \quad \text{and} \quad \eta_1 = \eta_1^0 \sin\omega(t - x/u_2),$$

and thus $\overline{\dot{\xi}_1^2} = \tfrac{1}{2}(\omega\xi_1^0)^2$ and $\overline{\dot{\eta}_1^2} = \tfrac{1}{2}(\omega\eta_1^0)^2$.

Also $\overline{(\partial\xi_1/\partial x)^2} = \tfrac{1}{2}(\omega\xi_1^0/u_1)^2 = \overline{\dot{\xi}_1^2}/u_1^2$

and $\overline{(\partial\eta_1/\partial x)^2} = \tfrac{1}{2}(\omega\eta_1^0/u_2)^2 = \overline{\dot{\eta}_1^2}/u_2^2$.

Using the values deduced for u_1 and u_2 in the last section, we see at once that each of the two brackets forming the Lagrangian density vanishes when averaged over time. Since by Fourier analysis any plane wave may be split into a series of pure sine waves, this result holds for all plane waves. The result is thus a special case of the general proposition of the extremum of the time-integral of the Lagrangian function.

§ 11. THE SHAPE OF THE VELOCITY–TEMPERATURE CURVE FOR SECOND SOUND ON THE BASIS OF LANDAU'S THEORY

Landau (1941) used the formula $u_2^2 = TS^2\rho_s/C\rho_n$ for the velocity of second sound with calculated as opposed to experimental (or empirical) values of the thermodynamical quantities involved. These calculated expressions contain at least two adjustable constants (arising in the roton contributions in Landau's terminology: in the case of a sharp energy gap as considered by Landau in his 1941 paper, just two constants, Δ and μ, are required, Δ giving the magnitude of the energy gap and μ the effective mass of the roton). The statement made originally by Lifshitz (1944) that Landau's expressions give for the velocity of second sound a monotonic decreasing function of temperature is only true for certain ranges of values of Δ and μ. It will now be shown that if the rôle of phonons is ignored (using Landau's terminology: phonons are Debye waves), the velocity at about 1·5° K. should rise with temperature, and this value should be almost independent of the particular value of Δ chosen. It is therefore possible to choose μ and Δ in such a way that the roton contributions overweigh those of the phonon contributions and yet still give the correct magnitude for the velocity of the sound. In this way it is possible to obtain a dependence of velocity of second sound on temperature far from monotonic, having a new maximum between the λ-point and the maximum at absolute zero. It is impossible as yet to deduce Δ except crudely, as inertia (Andronikashvilli 1946) and specific heat experiments are not sufficiently accurate; as has been pointed out already, the velocity of second sound is largely independent of Δ in the range of temperatures so far employed for these experiments.

To deduce this from Landau's theory (1941), we start with the formula $u_2^2 = TS^2\rho_s/C\rho_n$. Now at about 1·5° K. $\rho_s \simeq \rho$, the factor ρ_s/ρ_n merely giving the sharp decrease of velocity near the λ-point. From Landau's paper we may deduce

$$(\rho_n)_{\text{roton}} \propto u^{5/2}T^{3/2}e^{-\Delta/KT},$$

$$(C)_{\text{roton}} \propto u^{3/2}T^{-\frac{1}{2}}\Delta^2 e^{-\Delta/KT}\left\{1 + \frac{3KT}{\Delta} + \frac{15}{4}\left(\frac{KT}{\Delta}\right)^2\right\},$$

$$(S)_{\text{roton}} \propto u^{3/2}T^{\frac{1}{2}}\Delta e^{-\Delta/KT}\{1 + 5KT/2\Delta\}.$$

Thus, when only roton contributions are of importance,

$$u_2^2 = \frac{TS^2\rho_s}{C\rho_n} \propto \frac{T}{\mu}\left\{1+\frac{5KT}{2\Delta}\right\}^2 \Big/ \left\{1+\frac{3KT}{\Delta}+\frac{15}{4}\left(\frac{KT}{\Delta}\right)^2\right\}.$$

Since $KT \ll \Delta$ at $T \sim 1{\cdot}5°$ K., we see that under these conditions u_2 must rise with temperature.

Landau's results that the maximum value of the velocity is at absolute zero, and is equal to $u_1/\sqrt{3}$, will hold provided (i) any roton contributions have vanished due to the energy gap in the roton spectrum, and (ii) one part of the liquid has entropy zero and the other has an entropy due to Debye waves (phonons).

The temperature, according to Landau's theory, at which the rapid rise in the velocity of second sound takes place (at these temperatures close to the absolute zero) is governed chiefly by the value of Δ, which, as we have seen, is very difficult to derive experimentally from other sources. A value as low as about half a degree would be quite a possibility without contradicting present experimental evidence. Even this assumes the existence of a well-defined energy gap Δ. In this latter connection, a recent note by Landau (1947) is interesting.

§ 12. A CRITICISM OF TISZA'S THEORY

We only consider those aspects of Tisza's paper (1947) which directly concern the problem of the propagation of second sound.

Firstly, there is the problem as to whether it is possible for Debye waves to pass through the superfluid. Unlike Landau (1941), who includes phonons (i.e. Debye effects) as part of the normal fluid, Tisza considers that Debye waves are associated with the liquid as a whole, and that their presence will cause an attenuation of second sound. However, in his quantitative theory Tisza neglects their effect. If Tisza's viewpoint is correct, his theory must break down at temperatures below about 1° K., and his derivation of the zero value of the velocity of second sound at absolute zero (p. 852) and equations (11) and (12)—$d\rho_n/\rho_n = dS_n/S_n$; $\rho_n/\rho = S_n/S_0$—appear therefore to be invalid. Further, in the deduction of these equations Tisza considers the entropy as trapped in a cell possessing a membrane permeable only to superfluid. The argument of p. 845 is only true provided no heat is supplied from outside, since any heat supplied would create more normal fluid and, therefore, more entropy in the cell. Under the adiabatic conditions the equations are thus correct, but equation (12) will merely tell us the temperature in the cell, and will not, as Tisza claims, apply to equilibrium values at any fixed temperature we like to name. Hence we do not agree that Tisza's equation (13) $\{\rho_n/\rho = (T/T_0)^r$, where T_0 is the temperature of the λ-point$\}$ follows from (8) $\{S_n/S_0 = (T/T_0)^r\}$, assuming that S_n and ρ_n refer to equilibrium values at any given temperature. Also it is not clear physically why there should exist any thermodynamic relationship between the equilibrium values of S_n and ρ_n for any given temperature; one would expect the actual relationship between them to be specific to the detailed mechanism involved.

The values which are required in the formula for the velocity of second sound are ordinary equilibrium values of S, C and ρ_n for a given temperature T'. These mean equilibrium values,† about which the adiabatic fluctuations comprising second sound occur, are not directly connected with the values relating to Tisza's

† Actually not all the variables in this formula for the velocity of second sound are independent. Since $C = T'(\partial S/\partial T)$ and $\rho_s = \rho - \rho_n$, we may transform the formula to become $u_2^2 = (1-\rho/\rho_n)/\partial(1/S)/\partial T$. Thus the mean equilibrium values required for any given temperature are those of ρ_n and $\partial(1/S)/\partial T$.

"isopycnic" compression experiment. Thus it is incorrect to apply equation (12) (as Tisza does) to obtain the value of the velocity of second sound at $T=0$, and it is also irrelevant that Landau's phonon equations violate this equation, as Tisza points out on p. 852 of his paper. It must be added that the relationships derived by Landau (1941) for his roton effects also conflict with Tisza's equation (12), as may be seen from the values given in § 11 of this paper. (The entropy and normal density of an ideal Bose-Einstein gas do happen to obey Tisza's relationship, but it is well known that the values calculated on this very crude model are quite inaccurate, the entropy being given as proportional to $T^{1\cdot5}$, as opposed to the experimental relation $S \propto T^{5\cdot5}$.)

The present situation with respect to the temperature variation of the velocity of second sound might be summed up by saying that Landau and Tisza now agree on the formula $u_2^2 = TS^2\rho_s/C\rho_n$, although Tisza believes that the value of S inserted in this formula should refer not to the total entropy, but to this value less the Debye contribution. Tisza also considers that the formula will become inadequate below about 1° K., where these Debye effects become important. This author relies on empirical formulae for the entropy, with the constants determined from Kapitza's experimental values (Kapitza 1941): he derives his values of ρ_n from this entropy formula by the application of his relationship $\rho_n/\rho = S_n/S_0$, which, as we have shown above, has not the thermodynamic justification he has claimed. Landau goes further and attempts a theory involving two parameters, Δ and μ, capable of giving the values of S, C and ρ_n required, Δ and μ being determined from Keesom's values of the specific heat (Keesom 1942).

ACKNOWLEDGMENTS

The author would like to record his acknowledgment of the many helpful discussions he has had with Professor N. F. Mott of this Laboratory, Professor J. F. Allen of the University of St. Andrews, and colleagues of the Royal Society Mond Laboratory, Cambridge. He is also indebted to the Department of Scientific and Industrial Research for a grant.

REFERENCES

ANDRONIKASHVILLI, 1946, *J. Phys. U.S.S.R.*, **10**, 201.
GOGATE, D. V., and PATHAK, P. D., 1947, *Proc. Phys. Soc.*, **59**, 457.
KAPITZA, P., 1941, *J. Phys. U.S.S.R.*, **4**, 181; **5**, 59.
KEESOM, W. H., 1942, *Helium* (Elsevier Publishing Co.), p. 217.
KEESOM, W. H., and KEESOM, A. P., 1936, *Physica*, **3**, 359.
LANDAU, L., 1941, *J. Phys. U.S.S.R.*, **5**, 71; 1944, *Ibid.*, **8**, 1; 1947, *Ibid.*, **11**, 91.
LANDAU, L., and LIFSHITZ, E., 1938, *Statistical Physics* (Oxford : University Press), p. 114.
LANE, C. T., *et al.*, 1947, *Phys. Rev.*, **71**, 600.
LIFSHITZ, E., 1944, *J. Phys. U.S.S.R.*, **8**, 110.
LONDON, F., 1938, *Phys. Rev.*, **54**, 947 ; 1939, *J. Phys. Chem.*, **43**, 49.
MEYER, L., and BAND, W., 1947, *Phys. Rev.*, **71**, 828.
PESHKOV, V. L., 1946, *J. Phys. U.S.S.R.*, **10**, 389.
RAYLEIGH, Lord, 1926, *Theory of Sound*, 2nd Edn. (London : Macmillan), p. 318.
TISZA, L., 1938, *C. R. Acad. Sci., Paris*, **207**, 1035, 1186; 1940, *J. Phys. Radium*, **1**, 165, 350; 1947, *Phys. Rev.*, **72**, 838.

Electronic Structure of Fulvene

By C. A. COULSON*, D. P. CRAIG† and A. MACCOLL†
* Kings' College, London † University College, London

MS. received 14 *December* 1947

It is a well known fact that in dealing with most common conjugated and aromatic systems, such as benzene or the polyenes, the method of molecular orbitals (m.o.) and the method of valence-bond structures (v.b.) agree remarkably closely. This agreement includes bond orders, bond lengths, free valence and resonance energy. However, during a study of the fulvene molecule $C_5H_4.CH_2$ we have found a large and serious discrepancy between the predictions of the two methods of approximation. These discrepancies are large enough to be used to distinguish between the reliability of the methods by an appeal to experiment. Unfortunately the structural data are not yet available. But in view of the desirability of obtaining as many comparisons and tests as possible between the theories, it seems good to place on record the calculations that we have made, in the hope that they may stimulate experimental work with which to compare them. The calculations themselves are of quite standard pattern and we shall not therefore describe them in any detail.

In the v.b. calculations, resonance is included between five canonical structures. In the first place these are taken to be A –E.

Figure 1.

The normalized wave function for the ground state is

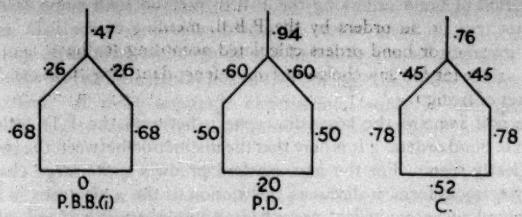
$$\Psi = 0{\cdot}6324\,\psi_A - 0{\cdot}2109\,(\psi_B + \psi_C) + 0{\cdot}4217\,(\psi_D + \psi_E). \quad \ldots\ldots(1)$$

The resonance energy, $0{\cdot}5\alpha$, has already been given by Wheland (1941). Bond orders may easily be inferred by the method of weighted contributions, as suggested by Pauling, Brockway and Beach (1935), or by the method of Penney (1937), using the Dirac coupling formula. The values are shown in figure 2, with the labels P.B.B. (i) and P.D. respectively.

The m.o. calculation of bond orders are made by the method of Coulson (1939). These values are shown in the figure with the label C. The resonance energy, $1{\cdot}47\beta$, has already been given by Wheland (1941).

Several comments suggest themselves at once from a study of figure 2. Firstly, the differences in bond order are much greater than is usually found, and even the sequence of orders is not preserved in any two of the diagrams. So far as the P.B.B. calculations are concerned this may be attributed to the impossibility of making any really satisfying choice of canonical structures. An objection

to the choice A–E is that none of them shows the bottom bond as double, so that in the P.B.B. method it cannot have any double bond character. The only way to avoid this is to make a different choice of fundamental structures. Any independent set of five will do, even if, in the language of Pauling, they are not canonical. This suggests that we introduce F and G. The appropriate modification of the wave function (equation (1)) is easily made for any independent



Valence-bond Molecular-orbital

Figure 2. Calculated bond orders.

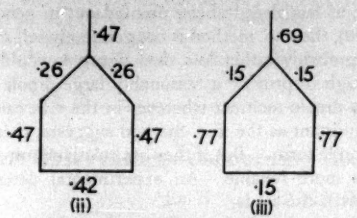


Figure 3. Alternative P.B.B. calculations of bond order.

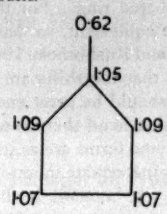


Figure 4. Charge migration according to m.o. theory. Figures show total mobile charge on each atom.

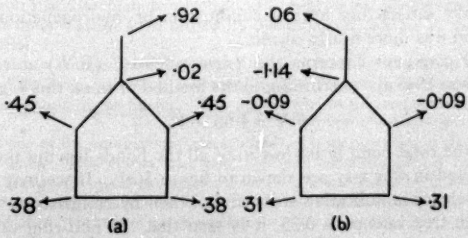


Figure 5. Free valence.
(*a*) Molecular-orbitals F.V.$=4{\cdot}68-\mathcal{N}$.
(*b*) Valence-bond, relative to an ethylenic carbon, so that F.V.$=4-\mathcal{N}$.

set, using the "un-crossing rule". In this manner we have used three reasonable sets, viz. (i) A–E, (ii) A, B, C, F and G, (iii) A, D, E, F, and G. Of these only (i) is a canonical set, but (iii) has the advantage of not using structures with more than one formal bond. The results of our calculations in case (i) are already given in figure 2 : the results for cases (ii) and (iii) are shown in figure 3. Even

the most cursory glance at these diagrams shows conclusively that there is not even the remotest agreement between them. It is quite evident that in molecules of this kind, in which "chemical intuition" does not suggest one single sensible choice of structures, the whole idea of weighted contributions is inadequate. Indeed, it suggests that even in other molecules such as anthracene or chrysene, although the choice of structures is reasonably straightforward, we must treat all calculation of bond orders by the P.B.B. method with some reserve. The same is not true for bond orders calculated according to the P.D. method, for these are invariant for any choice of five independent structures, and we obtain the same values (figure 2) whether we use the set (i), (ii), (iii) or any other independent set of five.

There still remains the large discrepancy between the P.D. valence bond and the m.o. bond orders. It is here that the distinction between the two methods is most clearly seen. For the m.o. method predicts quite large charge shifts. In its conventional form, it discusses the motion of the π electrons in a potential field supposed to be equivalent around each nucleus. But if we calculate the resulting charge distribution we find (figure 4) a large drift of electrons into the five-membered ring. For that reason the potential field around each nucleus cannot be equivalent, so that, as has already been pointed out in general terms (Coulson and Rushbrooke 1940), the m.o. method is not genuinely self-consistent. It is true that the shifts are probably rather less than figure 4 would indicate, but they should be great enough to provide a reasonably large dipole moment. On the other hand there is no dipole moment whatever in the v.b. calculations. Thus, if ionic terms are as important as the m.o. method suggests, then the v.b. method is inadequate in its simple form. But if they are relatively unimportant, then the v.b. method is the more reliable. An experimental determination of the dipole moment would settle this issue.

A further discrepancy between the two methods is found in the calculated energies. For a large number of molecules (Wheland 1941), the ratio of the resonance integral (β) to the exchange integral (α) required to give identical resonance energies has the value $\beta/\alpha = 0{\cdot}54$. But in the case of fulvene it is 0·34. A really satisfactory heat of combustion or hydrogenation could settle which method was more nearly correct.

Our final comment concerns the "free valence" (F.V.) at any particular centre (Coulson 1948 a). According to the method of m.o., this is given by

$$\text{F.V.} = 4{\cdot}68 - \mathcal{N}, \qquad \ldots\ldots(2)$$

where $\mathcal{N}$ is the total bond order ($\sigma+\pi$) of all the bonds leaving the given atom. Values obtained in this way are shown in figure 5(*a*). If we may suppose that a high free valence is indicative of reactivity, then by comparison with benzene, for which the free valence is 0·35, it is seen that the non-ring carbon atom is extremely reactive. Some measure of polymerization, with or without dismutation, would therefore be anticipated. This end carbon atom may be compared with the corresponding atom of benzyl, for which the free valence (Coulson 1948 b) is only slighter greater, at 1·04 instead of 0·92.

In cases where the choice of canonical structures in the resonance method is fairly evident, it is possible (Pullman 1947) to use the method of weighted contributions from each member of the set to estimate free valence, in a manner similar to that previously described for bond orders. But in the case of fulvene,

the reasons which we have given for disregarding the bond orders thus calculated, apply equally to the free valence, and we shall not therefore use this method. A similar objection does not obtain (Moffitt 1948) with the P.D. type of calculation, for which a description of the free valence may be given, parallel to that implied in equation (2) above for the molecular orbital theory. But since the greatest possible value of total bond order at a carbon atom (corresponding to 4·68 in m.o. theory) has not yet been settled for v.b. theory, our best way is to measure free valence relative to an ethylenic carbon. The resulting values are shown in figure 5 (*b*). The high reactivity of the end carbon compared with the other carbons is now not at all noticeable and the difference between figures 5 (*a*) and 5 (*b*) is marked. A more adequate experimental study of the chemistry of fulvene should help to resolve the question which of 5 (*a*) and 5 (*b*) better represents the molecule.

REFERENCES

COULSON, C. A., 1939, *Proc. Roy. Soc.* A, **169**, 413 ; 1948 a, *Trans. Faraday Soc.*, **34**, 123 ; 1948 b, *Ibid.*, " Discussion on Free Radicals ".
COULSON, C. A., and RUSHBROOKE, G. S., 1940, *Proc. Camb. Phil. Soc.*, **36**, 193.
MOFFITT, W. E., 1948, *Trans. Faraday Soc.*, in the press.
PAULING, L., BROCKWAY, L. O., and BEACH, J. Y., 1935, *J. Amer. Chem. Soc.*, **57**, 2705.
PENNEY, W. G., 1937, *Proc. Roy. Soc.* A, **158**, 306.
PULLMAN, A., 1947, *Ann. Chim.*, **2**, 1.
WHELAND, G. W., 1941, *J. Amer. Chem. Soc.*, **63**, 2025.

Some Observations concerning the Energy of Nuclei

By E. GLUECKAUF
Atomic Energy Research Establishment, Harwell, Didcot

MS. received 16 *December* 1947

ABSTRACT. The α-binding energy as a function of the number of neutrons and protons is shown to exhibit sharp fluctuations which occur at more or less regular intervals. There is evidence that these " kinks " are caused by discontinuous changes in the density and radius of the nuclei.

It is shown that a mathematical connection exists between " kinks " in the α-binding energy curve and the irregularities of the β-stability line (Gamow valley), so that both must occur in the same region of the nuclei. This applies also to the irregularities in the mass-defect curve.

A study is made of the energies of β-decay (ΔM) of isobaric nuclei in the region of the radioactive elements. This shows that nuclei of odd mass-number A have a slightly higher binding energy for even proton number Z (a lower mass by 0·2 MEV.) than for odd Z. The effect, though well known for nuclei of even mass-numbers, has not before been observed for odd ones, where it is much smaller.

§1. REGULARITIES IN THE α-STABILITY OF NUCLEI

THE purpose of this section is to draw attention to certain regular variations in the α-binding energy of nuclei which are superimposed on the general trend of decreasing α-stability with increasing mass and charge. Being related to the gradient of the mass-defect curve, the α-stability is better suited than the mass defect itself for exhibiting irregularities which point to structural features in nuclei.

We shall employ the usual terminology, denoting the atomic weight by A, the atomic number by Z, the number of neutrons by $N(=A-Z)$, and the isotopic number by $I\,(=N-Z=A-2Z)$.

(i) *Data for heavy nuclei*

One of the well known features of all the radioactive series is the extremely marked change in the α-energies between the isotopes of atomic number $Z=82$,

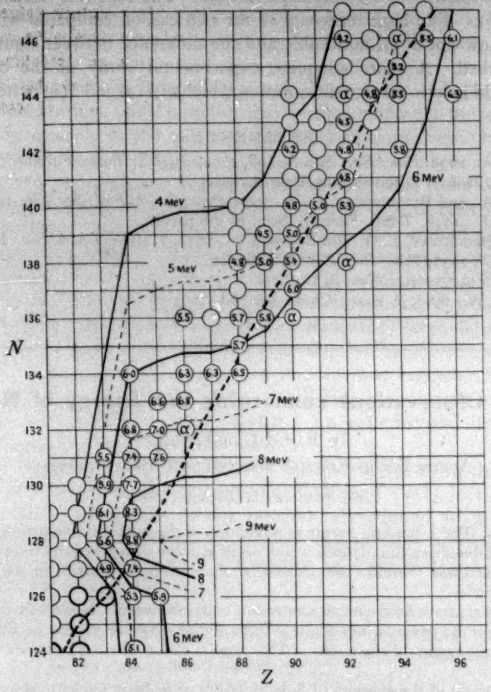


Figure 1. α-energies of large nuclei (MeV.).
Heavy circle: Stable nuclei. Light circle: β-active nuclei.
Light circle surrounding figures: α-active nuclei with or without β-decay.

which are, within the limits of experimental accuracy, completely stable against α-decay, and those of atomic number $Z=84$, which have the most short-lived bodies and emit the most powerful α-particles. The average rate of change of the α-energy E_α in this region is $(\Delta E_\alpha/\Delta Z)_{N=\text{const.}}=2{\cdot}1$ MeV. per proton. After element 84, the α-energy increases only slightly, $[\Delta E_\alpha/\Delta Z]_N$, varying from 0.25 to 0 to 0.5 MeV. per proton.

This striking discontinuity at the proton number $Z=84$ has given rise to much valuable discussion (notably by Feather 1945), during which, however,

it has apparently been overlooked that a discontinuity of similar magnitude and importance exists also at the neutron number $N=128$. If we plot the energy of the emerging α-particles against the neutron number N and draw lines of constant Z, we get a sharp rise in E_α between $N=126$ and $N=128$ with a gradient of $[\Delta E_\alpha/\Delta N]_{Z=\text{const.}}=1{\cdot}9$ MeV. per neutron, followed after $N=128$ by a constant decrease $[\Delta E_\alpha/\Delta N]_Z=-0{\cdot}5$ MeV., which lessens gradually after $N=138$ until it reaches a value of $-0{\cdot}1$ MeV. at $N=144$.

No systematic diagram of the α-energies is therefore quite complete which does not show this dual system of discontinuities at $Z=84$ and at $N=128$. It would appear that the best way of representing the facts is shown in figure 1. This represents a map of nuclei, using Z and N as abscissa and ordinate respectively, and the α-energy is indicated for every nuclear species. We can then draw contour-lines of constant α-energy, treating that quantity as if it were a continuous variable of Z and N. The diagram is ideally suited for the interpolation and extrapolation of unknown α- and β-energies.

To give an example, it is possible to predict with a fair degree of certainty that the only β-stable α-emitter of element 97 will have mass 247, an E_α of 6·0 to 6·1 MeV. and a half-life of the order of one year. As the highest mass number so far reported in a trans-uranic element is 242, it is thus not likely that this new species will be discovered in the near future.

Though in figure 1 the discontinuities at $Z=84$ and $N=128$ are very clearly noticeable, the general impression is that these discontinuities are not as important as they appear if Z or N alone is chosen as variable. It rather appears that the lines of constant E_α form an irregular wave, with its axis somewhat steeper than the band of the known α-emitting nuclei which follow approximately the line of maximum β-stability (dotted line). The question immediately arises whether this E_α "wave" is confined to the radioactive elements or whether it can be traced also in the binding energies of the α-stable nuclei.

(ii) *Data for light nuclei*

For the lightest elements up to a mass-number A of about 40, the nuclear masses M are well enough known to permit a fairly accurate calculation of the binding energy E'_α for the last α-particle (which is the α-energy with opposite sign), according to

$$E'_\alpha={}^{A-4}_{Z-2}M+{}^{4}_{2}M-{}^{A}_{Z}M.$$

With reduced accuracy these calculations can be extended to $A=60$. The mass data have been taken from an isotope chart compiled by Perlman, Seaborg and Segrè in May 1947, and the α-energies are given in milli-mass-units (m.m.u.).

Plotting again lines of constant E'_α in a $(Z-N)$-diagram (figure 2), we find that, here too, the E_α-lines perform irregular but more or less parallel wave motions, with the wave-axis approximately parallel to the Weizsäcker line of stable elements. The bulges near $A=19$, 39 and 59 represent minima of the α-binding energy differing by as much as 8 m.m.u. from the normal level of E'_α. The gradients $[\Delta E_\alpha/\Delta Z]_N$ and $[\Delta E_\alpha/\Delta N]_Z$ near the minima of E'_α are actually of the same order as those observed with the radioactive elements near $Z=83$, i.e. about 2 MeV. per added nucleon.

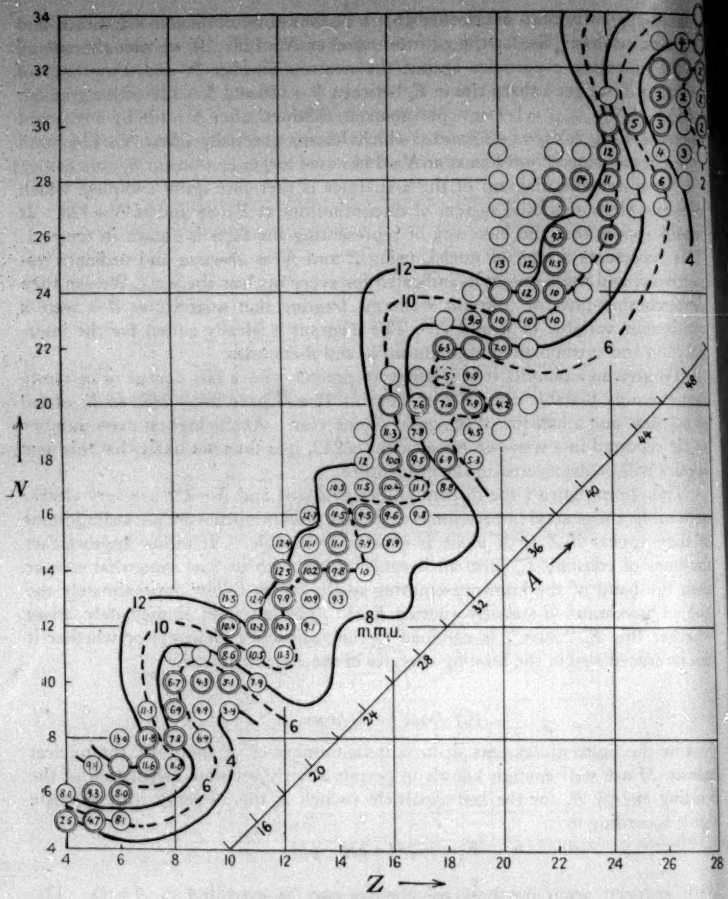


Figure 2. α-binding energies of small nuclei (in m.m.u.).

Double circle: Stable isotopes.
Single circle: Isotopes unstable against β^-, β^+ or K-decay.

(iii) *Discussion*

There is a considerable temptation to interpret these recurring fluctuations of the α-binding energy, which must be expected also near $A = 90$ and 140 (viz. the α-activity of ^{148}Sm), in terms of a "shell" model, though this need not necessarily be of the types considered by Elsasser (1933, 1934) and by Margenau (1934). With a shell structure it appears plausible that the addition of both neutrons and protons to a complete shell has at first an unstabilizing effect.

The α-stability would decrease until the newly added nucleons are numerous enough to interact with each other with a sufficiently large binding energy. From this point onwards the α-stability should increase rapidly to its normal value and remain there—with minor fluctuations—until the shell is filled, after which the process repeats itself.

For small nuclei the electrostatic forces acting on a proton are small, and there is thus not much difference between the addition of a neutron or of a proton. The maxima and minima of E_α should, therefore, be reached at a given number of nucleons independent of whether N or Z or I is kept constant. This is borne out well in figure 2. A similar picture can be applied to the elements of high atomic weight, but here the electrostatic effect of the high nuclear charge, resulting in a tilt of the curve, comes significantly into play; consequently minima of the binding energy (i.e. maxima of the α-energy) are obtained only for lines of constant Z or I, and no longer for constant N.

While this "shell" model must be considered only as tentative, it should be pointed out that it does explain, at least qualitatively, the very puzzling feature that, at the proton numbers $Z=8$, 9, 17, 18, and 26, the addition of successive neutrons reduces the α-binding energy, and that, at the neutron numbers $N=10$, 11 and others, the addition of successive protons—in spite of their positive charge—increases the binding energy of the last α-particle.

(iv) *The case of argon* 39

The α-binding of ^{39}A, which in figure 2 is given as approximately 5 m.m.u., has been derived in the following way. It follows from the mass difference of $^{35}_{16}$S (34·9788) and $^{43}_{20}$Ca (43·9723) that $(^{43}_{20}E'_\alpha + {}^{39}_{18}E'_\alpha) = 14{\cdot}3$ m.m.u.

From its position in figure 2 one would expect $^{43}_{20}E'_\alpha$ to be about 9 m.m.u., which leaves for $^{39}_{18}E'_\alpha$ only about 5·3 m.m.u. But even if this value of E'_α for ^{39}A is not accepted, it is apparent from the β-energy of $^{35}_{16}$S of only 0·11 m.m.u., and from the fact that ^{39}A must be unstable, that $^{39}_{18}E'_\alpha \leqslant {}^{39}_{19}E'_\alpha$. This fact is of considerable interest, since all theories of the nuclear masses based on a nuclear radius r which grows smoothly with increasing mass ($r \propto A^{\frac{1}{3}}$), require that isobars have an α-binding energy which increases with the isotopic number I (Weizsäcker (1935), Wigner (1937, 1939), or Feenberg (1947)). The case of ^{39}A obviously contradicts this assumption of a smoothly growing radius and requires that a sharp expansion of the nuclear radius takes place in the region of ^{39}A.

To a minor extent this phenomenon occurs also at the other minimum, at $A=17$ and 19. Here the differences between the E'_αs of isobars for $I=+1$ and $I=-1$ are only about half of the average $\Delta E'_\alpha$ between $A=21$ and 37, while according to all the theories they should be larger than the $\Delta E'_\alpha$s of subsequent nuclei ($\Delta E'_\alpha$ approx. $\propto A^{-\frac{1}{3}}$).

In addition to this it is known from the application of the Gamow equation to the data of the decay constants and α-energies of the radioactive elements, that a sharp expansion of the nuclear radius takes place in the region of the elements between $Z=81$ and 83 (Feather 1945), which is exactly where the biggest rise in the α-energies takes place. We therefore conclude that the minima of the α-binding energy shown in figure 2 are also caused by a sudden expansion of the nucleus with increasing mass. While the nuclear radius seems to grow, on the average, with $A^{\frac{1}{3}}$, the possibility must be considered that

it grows only little, if at all, in the interval between two "bulges" of the α-binding energy. This means that the density of the nucleus must be expected to undergo regular fluctuations, minima of the density coinciding with minima of α-stability.

There is some other evidence which supports the conception that the fluctuations in the α-binding energy are not a purely energetic feature but are due to changes of the internal structure.

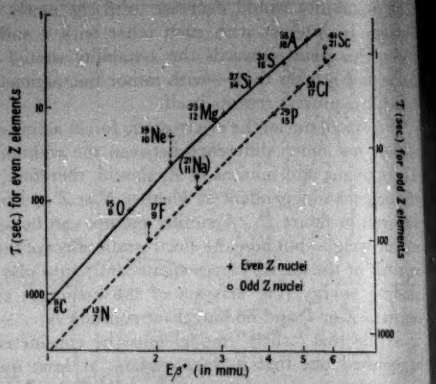


Figure 3. Deviations from Sargent curve near $A=19$ and 39. (Logarithm of half-life plotted against ln E_{β^+}).

If the Sargent curve of ln $E_{\beta+}$ against ln τ is plotted for the β^+-transitions from $I=-1$ to $I=+1$ (figure 3), there is a marked deviation from the otherwise smooth line in the region of $A=19$. (Such a deviation can also be detected in the function $F\tau$ calculated by Konopinski (1943), which represents essentially the same feature.) It follows that the decay of ^{19}Ne is 2·2 times faster than one would expect from its β^+-decay energy. There is also a small displacement in the same direction for ^{17}F, ^{41}Sc and most likely also for ^{21}Na.

A similar phenomenon occurs for β^- transitions from $I=+3$ to $I=+1$. In the case of ^{19}O the β-decay is about three times slower than one would expect from the extrapolation of the Sargent curve drawn through the points of ^{23}Ne and ^{27}Mg. In either case the nuclei near the minimum of E'_α behave as if the nucleus with the higher isotopic number had a smaller mass than it actually has. The effect reminds us of the similar deviations in the α-decay constants of element 83 in the Geiger-Nuttall plot, which are caused by a comparatively large change of the nuclear radius and density.

§2. THE INTERDEPENDENCE OF MASS DEFECT, α- AND β-STABILITY

Feenberg (1947) has shown by semi-empirical considerations that a relationship exists between the mass-defect curve and the β-stability curve of nuclei, and that the irregularities of the one correspond to deviations of the other curve at the same values of the mass-number A. This follows directly—as was shown independently by the author—from the fact that the sum of the α- and β-energies in a cycle must be zero.

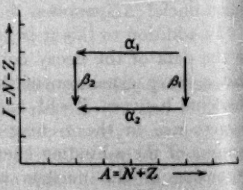


Figure 4. α and β decay cycle.

It may be seen from figure 4 that

$$4[\Delta E_\beta/\Delta A]_{I\ \text{const.}} - 2[\Delta E_\alpha/\Delta I]_{A\ \text{const.}} = 0. \quad \ldots\ldots(1)$$

This can be transformed into

$$\left[\frac{\Delta I}{\Delta A}\right]_{E_\beta=\text{const.}=0} = -\left[\frac{\Delta E_\alpha}{\Delta I}\right]_A \Big/ 2\left[\frac{\Delta E_\beta}{\Delta I}\right]_A = \left[\frac{\Delta E'_\alpha}{\Delta I}\right]_A \Big/ 2\left[\frac{\Delta E_\beta}{\Delta I}\right]_A, \quad \ldots\ldots(2)$$

where E'_α is the binding energy of the last α-particle in the nucleus, and where $[\Delta I/\Delta A]_{E_\beta=0}$ is the gradient of the line of maximum β-stability, i.e. the bottom of the Gamow valley. $[\Delta E_\beta/\Delta I]_A = [\Delta^2 M/(\Delta I)^2]_A$ is known to be a comparatively smooth and very slowly changing function of A for nuclei of odd mass-number A, which is always positive.

Normally E'_α increases with increasing isotopic number and $I^0 = [I]_{E_\beta=0}$ increases with increasing A.

Equation (2) thus requires that in the region of the well known kinks of the β-stability line, where I^0 decreases with increasing A, the α-binding energy E'_α must also decrease with increasing isotopic number I. Normally E'_α does not change very much for neighbouring nuclei with the same isotopic number, but in order to fulfil this condition it is necessary that E'_α undergo a considerable change within a series of nuclei of the same value of I.

The mass-defect D is related to the α-binding energy by the equation

$$[\Delta D/\Delta A]_{I=\text{const.}} = \tfrac{1}{4}(D_\alpha - E'_\alpha), \quad \ldots\ldots(3)$$

where D_α is the mass-defect of ${}^4_2\text{He}$. Equation (3) shows that the mass-defect curve is approximately linear only when E'_α is reasonably constant, and that a marked fluctuation of E'_α must cause a similar deviation of the mass-defect curve from near-linearity.

It thus follows from equations (2) and (3) that in the regions where the β-stability line shows pronounced kinks, as, for example, near the elements ${}_{18}\text{A}$, ${}_{43}\text{Tc}$, ${}_{58}\text{Ce}$ and ${}_{51}\text{Il}$, there must be equivalent kinks in the α-stability lines and in the mass-defect curves. Vice versa, in such regions where the α-stability is known to show marked fluctuations, e.g. near ${}_{19}\text{F}$ and ${}_{83}\text{Bi}$, similar deviations must also be expected for the β-stability line and, obviously, for the mass-defect curve.

§ 3. THE ISOBARIC CROSS-SECTION THROUGH THE NUCLEAR MASS SURFACE IN THE REGION OF THE HEAVY ELEMENTS

The preceding study of the α-energies of the radioactive nuclei of large mass has resulted in a diagram (figure 1) which renders possible the interpolation of unknown α-energies with an accuracy of approximately $\pm 0{\cdot}1$ MeV., and a similar, if not better, accuracy for the differences of E_α for isobars (nuclei of constant mass). This makes it possible to calculate also the β-energies (E_β) of unknown nuclei from known β-energies by the relation

$${}^A_Z E_\beta = {}^{A-4}_{Z-2}E_\beta + {}^A_Z E_\alpha - {}^A_{Z+1}E_\alpha$$

(E_β stands here for the total energy change in a β-transition equal to ΔM).

The accuracy of the β-energies so obtained depends on the number S of such steps required to obtain the E_β in question, but it should be better than $\pm 0{\cdot}1\,S$ MeV. If it were a purely statistical error it would be $0{\cdot}1\sqrt{S}$ MeV.

This procedure makes it possible to calculate the energies of β-decay (ΔM) of a much larger number of isobars than is obtainable in any other region of the

system of nuclei, not excluding the region of the fission products, and to investigate how far our assumptions concerning nuclear masses are correct, as the β-energy is the gradient of the nuclear mass surface along a line of constant mass-number.

Of particular interest are the isobars of odd mass number A for which the current theories predict a linear variation of ΔM with the isotopic number I.

The calculations have been carried out for the largest and the smallest mass-numbers of the two types of odd A, the mass differences of which can be obtained safely, i.e. 233 and 221 for the series $(4n+1)$ and 239 and 219 for the series $(4n+3)$. These two extremes give a very similar picture and, whilst in the one case the small values of Z and in the other case the large values of Z show the greatest uncertainty, one can be sure that the extrapolation has not caused any significant disturbance.

The results are shown in tables 1 (a) and (b).

The β-energies used for the extrapolations are (in MEV.)

for the $(4n+1)$ series :

$$^{233}_{90}\,1{\cdot}6,\quad ^{233}_{91}\,0{\cdot}8,\quad ^{225}_{88}\,{<}0{\cdot}05,\quad ^{213}_{83}\,{\sim}1{\cdot}3,\quad ^{209}_{82}\,0{\cdot}7;$$

for the $(4n+3)$ series :

$$^{239}_{92}\,1{\cdot}16,\quad ^{239}_{93}\,0{\cdot}67,\quad ^{231}_{90}\,{\sim}0{\cdot}2,\quad ^{227}_{89}\,0{\cdot}22,\quad ^{211}_{82}\,1{\cdot}4,\quad ^{207}_{81}\,1{\cdot}47.$$

Table 1(a). Type $4n+1$

$A=233$					$A=221$				
Z	I	ΔM	$\Delta^2 M$	Number of steps	Z	I	ΔM	$\Delta^2 M$	Number of steps
90	53				84	53			
		1·6		0			3·0		3 down
91	51		1·0		85	51		0·9	
		~0·6		0			2·1		3 down
92	49		2·4		86	49		1·5	
		−1·8		2 up			0·6		1 down
93	47		0·5		87	47		0·6	
		−2·3		5 up			~0		2 up
94	45		1·3		88	45		1·2	
		−3·6		6 up			−1·2		3 up
95	43				89	43			

Table 1(b). Type $4n+3$

$A=239$					$A=219$				
Z	I	ΔM	$\Delta^2 M$	Number of steps	Z	I	ΔM	$\Delta^2 M$	Number of steps
92	55				82	55			
		1·2		0			3·9		5 down
93	53		0·5		83	53		0·4	
		0·1		0			3·5		5 down
94	51		1·8		84	51		1·6	
		−1·1		2 up			1·9		3 down
95	49		0·9		85	49		0·5	
		−2·0		3 up			1·4		3 down
96	47		1·5		86	47		1·0	
		−3·5		7 up			0·4		2 up
97	45		0·8		87	45		0·9	
		−4·3		8 up			−0·5		3 up
98	43				88	43			

As Δ^2M within the same nuclear type is approximately the same for given values of I, the average of the values of Δ^2M are likely to give a more accurate picture; these are shown in table 2.

Table 2. Average Δ^2M

I	53	51	49	47	45
$A=4n+1$	—	0·95	1·95	0·55	1·25
$A=4n+3$	0·45	1·7	0·7	1·25	0·85

Near $A=229\pm10$.

It is apparent that Δ^2M is by no means constant, as required by the various theories (Weizsäcker 1935, Wigner 1937, 1939, Feenberg 1947), but that it varies in such a way that nuclei with an even value of Z always have a larger value of Δ^2M. This means that the mass-defect values of the isobars of odd mass-number plotted against I do not lie on a smooth parabola but on two parabolas, for odd and even values of Z respectively, the latter values being on the average 0·2 MeV. smaller than the former. The effect is similar to, but not quite so marked as, that of the nuclei of even mass-numbers, where the greater stability (lower mass) of nuclei with even values of Z is well known.

The trend in the variations of Δ^2M seems to be similar for the $(4n+1)$ and the $(4n+3)$ types of nuclei, and approximately the same values of Δ^2M occur in the case of the $(4n+3)$ types at values of I, which are two units higher than those for the $(4n+1)$ series. It would appear that this effect is another evidence of the formation inside the nucleus of "complexes" consisting of two protons and two neutrons which, in the presence of excess neutron pairs, would result in a somewhat greater stability for nuclei with all protons paired (even Z) than for nuclei with all neutrons paired (odd Z).

ACKNOWLEDGMENTS

The author wishes to thank Professor O. R. Frisch and Mr. J. M. C. Scott, A.E.R.E. Harwell, for many valuable discussions.

REFERENCES

Elsasser, J., 1933, *Phys. Radium*, **4**, 549 ; 1934, *Ibid.*, **5**, 389.
Feather, N., 1945, Department of Atomic Energy, Report *BR.* 640.
Feenberg, E., 1947, *Rev. Mod. Phys.*, **19**, 239.
Konopinski, E. J., 1943, *Rev. Mod. Phys.*, **15**, 209.
Margenau, H., 1934, *Phys. Rev.*, **46**, 613.
von Weiszäcker, C. F., 1935, *Z. Phys.*, **96**, 431.
Wigner, E., 1937, *Phys. Rev.*, **51**, 947 ; *see also* Barkas, W. H., 1939, *Ibid.*, **55**, 691.

The Temperature Effect on Cosmic-Ray Intensity and the Height of Meson Formation

By A. DUPERIER
Turner and Newall Research Fellow of the University of Manchester

Communicated by P. M. S. Blackett ; *MS. received* 10 *March* 1948

ABSTRACT. A study has been made of the results obtained in different parts of the world for the effects of ground temperature on cosmic-ray intensity. It is found that the great discrepancies between the values for the temperature coefficient as given by different observers, as well as the seasonal change of this coefficient, far from being a difficulty in Blackett's interpretation may be taken as further evidence of the instability of the meson. The 12-monthly variation in cosmic-ray intensity as obtained by Forbush can be also accounted for as an effect of meson decay if one assumes that the bulk of the penetrating component originates at 7·5 cm. Hg pressure-level. This assumption has been put to the test and the results obtained appear to confirm that the layer of meson formation lies at, or little above, the 7·5 cm. pressure-level, at the height of the maximum of the Pfotzer curve for the intensity of the total radiation.

§1. INTRODUCTION

In the last fifteen years, it has been reported by a number of investigators that the intensity of cosmic rays at constant atmospheric pressure shows from day to day and from month to month a variation inverse to that of the air temperature near the ground. It is well known that Blackett (1938) attempted to relate this temperature effect to the instability of the meson, and that in this way he explained the decrease of cosmic-ray intensity in warm weather as due to the greater height of the pressure level at which mesons originate. As Blackett pointed out, the coefficient of the temperature effect at the equator should be lower than at moderate latitudes since the incident primary rays are more energetic there.

However, when the experimental results obtained in different parts of the world are closely examined, several difficulties in this interpretation appear to arise. We have first the fact that the values of the temperature coefficient for stations at moderate latitudes, as obtained by different observers, show discrepancies among themselves which are much greater than would have been expected from the accuracy of the measurements. These discrepancies are particularly noticeable if we compare the values obtained by following one method of computation with those obtained by another. Some observers have correlated the monthly or seasonal means of cosmic-ray intensity with monthly or seasonal means of ground temperatures. Others have computed the coefficient for each month separately by using the daily means and then averaging the twelve coefficients. The former method appears to lead to values which are about twice as great as those obtained by the latter.

In the second place, when computed for each month separately, the temperature coefficient shows a most unexpected seasonal variation.

Forbush (1938) has shown that the annual variation in cosmic-ray intensity consists of two components—one of a world-wide character related to geomagnetic activity, and the other a 12-monthly wave inverse to that of the

temperature. There might be added as a further difficulty in the interpretation by meson decay the fact that the comparison of this 12-monthly wave with the annual change in ground temperature leads to a temperature coefficient which varies from one station to another.

In a former paper (Duperier 1941) the author showed, though in a manner which could not be considered as conclusive, that some at least of these difficulties would disappear if the mean temperature of the atmosphere, instead of the temperature near the ground, had been used for the correlation with cosmic-ray intensity.

The more complete information concerning the upper air, which has been gathered in the last few years in England by means of four daily radio-sondes from the same station, has made possible a further study of the extent to which the temperature effect may be related to the instability of the meson.

§2. THE TEMPERATURE EFFECT

In a previous publication (Duperier 1945) it was reported that the intensity of cosmic rays, as measured by a counter arrangement registering triple coincidences, appears to be closely related to the height of the 7·5 cm. Hg pressure-level. It was found that the correlation of cosmic-ray intensity at constant atmospheric pressure with the height of lower pressure-levels is rather low, though gradually increasing with height. When the 7·5 cm. pressure-level is chosen—the highest for which sufficient meteorological data was then available—this correlation reaches the values $-0{\cdot}67$. By admitting the principle of the instability of the meson and assuming that the bulk of the penetrating component originates at this pressure-level, the value 18·6 km. was obtained for L, the mean range of mesons.

Now, if δh represents the change in height corresponding to the variation δt of ground temperature, the coefficient α of the ground temperature effect will be given by the expression

$$\alpha = -(1/L)(\delta h/\delta t). \qquad \ldots\ldots(1)$$

As the heating of the atmosphere by conduction and convection, and so the expansion δh corresponding to a certain variation δt of the ground temperature, is the greater the longer the period during which this variation occurs, it is to be expected that the value of the ground temperature coefficient obtained by using variations from month to month will be greater than that obtained by the use of variations from day to day.

In the last few years it has been the practice in England to make radio-sonde observations every day at 00, 06, 12, 18 hours G.M.T. For the present study, the upper-air data obtained at Larkhill (about 100 km. to the south-west of London) during the period November 1943–October 1945 have been used, and the heights of the 7·5 cm. pressure-level determined.

The mean heights at 00, 06, 12, 18 hours for each season have been plotted in figure 1, together with the mean temperature near the ground.

It can be seen that, contrary to general belief, the mean height for each season changes quite appreciably with the time of day. As the number of days used for the computation of the means (more than 2000 in all) is sufficiently high practically to eliminate the effect of change in ground pressure, the variation in height of the 7·5 cm. pressure-level has to be ascribed to the thermal structure of the air below it.

To see, however, how closely the diurnal variation in height is related to the diurnal variation of ground temperature alone, departures from daily mean height for each hour have been correlated with similar departures of ground temperature. By doing this, the seasonal effect is obviously eliminated, and we find a value of $0{\cdot}92 \pm 0{\cdot}02$ for the correlation, and 18 ± 1 metres per °c. for b_1, the regression coefficient.

If we substitute b_1 for $\delta h/\delta t$ and 18,600 m. for L in (1), we have for the temperature coefficient $\alpha_1 = -0{\cdot}10\%$ of mean intensity per °c. This value can be compared with those given by Hess *et al.* (1940) and by Hogg (1947), since these

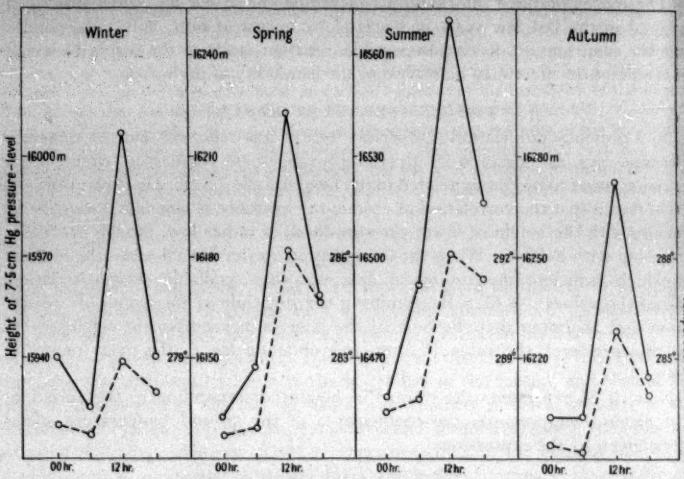


Figure 1. Height ——— Temperature - - - -

observers have followed a method of computation by means of which the seasonal effect is also eliminated. From series of observations of cosmic-ray ionization covering periods of several years, they obtained α_1 for each month separately by using daily means, and then averaging the twelve coefficients. As table 1 shows, the three values are the same.

Table 1

Locality	α_1 (%)	Observer
The Hafelekar (Austria)	−0·09 and −0·107	Hess *et al.* (1940)
Canberra (Australia)	−0·11	Hogg (1947)
London	−0·10	Duperier
Mean	−0·10	

Hess *et al.* and Hogg have also reported the existence of a marked seasonal change of the temperature effect. Hess has found that α_1 is nearly twice as great in winter as in summer. From Hogg's results it can be seen that the ratio of the two coefficients is 1·5.

That the change in height of the 7·5 cm. pressure-level is more pronounced in winter than in summer, compared with that in ground temperature, is already apparent in figure 1. From the regression coefficients for the winter and summer groups separately, the value 1·6 is obtained for the ratio α_{1W}/α_{1S}, which is quite consistent with those of Hess and Hogg.

If we correlate monthly averages of the means of the four heights and temperatures for each day (plotted in figure 2) instead of correlating daily departures, then we obtain for the correlation and regression coefficients the values $0{\cdot}98 \pm 0{\cdot}01$ and 45 ± 2 metres per °C. respectively. By applying the equation (1), we have now for the temperature coefficient $\alpha_2 = -45/18{,}600 = -0{\cdot}24\%$ per °C. which is about double the value of α_1 obtained before.

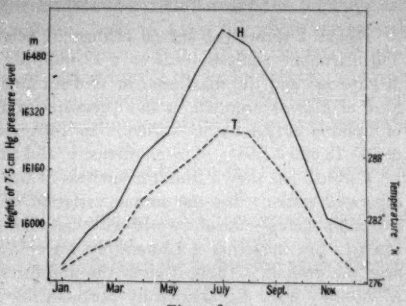


Figure 2.

The temperature effect obtained from monthly means should therefore be about twice as great as the value from daily means.

Table 2 gives the coefficients which have been obtained in different parts of the world by using monthly or seasonal means of cosmic-ray ionization and ground temperature.

Table 2

Locality	α_2 (%)	Observer
Voyage Vancouver (Canada) to Sydney (Australia)	−0·18	Compton and Turner (1937)
	−0·19	Gill (1939)
Cape Town	−0·12	Schonland *et al.* (1937)
Amsterdam	−0·21	Clay and Bruins (1939)
Mean	−0·18	

It can be seen that all the values are greater than those given in table 1 and that the mean is about twice as great, in agreement with the results from the radiosonde observations at Larkhill. The relatively small value for Cape Town, at 32·7° S. geomagnetic latitude, may be partly due to the primary radiation being more energetic there.

On the hypothesis of the instability of the meson, the use of daily means results in a smaller value for α on account of the lag in the warming of the atmosphere relatively to the warming of the ground, as previously stated. As the height of any pressure-level actually depends on the spatial temperature defined by $\int_1^z t(z)\,dz/z$ of the air below it, it is clear that had the cosmic-ray intensity been compared with this temperature, the result would have been the same independent of the length of the period over which the data are averaged.

In conclusion, the discrepancies between the temperature coefficients as obtained by different observers are due at least for the most part to the methods of computation having been different. Far from being a difficulty in the meson

decay interpretation, these discrepancies, together with the seasonal change of the ground temperature effect, may be taken as further evidence of the instability of the meson.

§ 3. THE 12-MONTHLY VARIATION

Figure 2 shows the annual change in height of the 7·5 cm. pressure-level. The harmonic analysis leads to a 12-monthly variation with an amplitude of 307 metres and the maximum in mid-summer. If again we assume that the bulk of mesons originate at this pressure-level we should have at the latitude of London a 12-monthly variation in cosmic-ray intensity with an amplitude of $307/18{,}600 = 1{\cdot}65\%$ of mean intensity, and the maximum in mid-winter.

Forbush has shown from the analysis of the observations made at five widely separated stations that the annual variation of cosmic-ray intensity consists of two components—one of a world-wide character corresponding with geomagnetic activity, and the other a 12-monthly wave inverse to that of the temperature near the ground. Table 3 gives the amplitudes and times of maxima of this 12-monthly wave as obtained by Forbush for The Hafelekar, Cheltenham and Christchurch, the latitudes of which are comparable to that of London.

Table 3

Station	Latitude	Geomag. lat.	12-monthly variation Ampl. (%)	Maximum
Cheltenham (U.S.A.)	38·7° N.	50·1° N.	1·6	January 19
Christchurch (N.Z.)	43·5° S.	48·0° S.	0·8	July 28
Hafelekar (Austria)	47·3° N.	48·4° N.	1·9	January 15
		Mean	1·43	
London (predicted)	51·5° N.	54·0° N.	1·65	January 18

It can be seen that the predicted amplitude and phase for London is in fairly good agreement with the mean value of the amplitudes and phases at the other three stations.*

As the upper air observations show that for stations at similar latitudes the range of the annual variation in atmospheric temperature, in contrast with the range of the variation of the ground temperature, is but little affected by the climate, it may be taken that the variation in height of the 7·5 cm. pressure-level in London is comparable with the mean of these variations at the other stations. If this is so, the agreement shown by the table may be taken as indicating strongly that the 12-monthly variation in cosmic-ray intensity is solely the effect of meson decay.

This conclusion seems to rule out the possibility of explaining the observed 12-monthly variation as due to the heliomagnetic field, as suggested by Vallarta and Godart (1939). If this field does not give rise to such a variation, the reason is perhaps that the numbers of positive and negative particles in the primary radiation are equal or, alternatively, it may be that the properties of the permanent magnetic field of the sun are such that this field does not change the intensity of the incoming radiation in the course of the year.

* Yearly variation not determinable by counter-apparatus.

§4. THE HEIGHT OF MESON FORMATION

We have seen in the preceding sections that the temperature effect as well as the 12-monthly variation of cosmic-ray intensity, as obtained in different parts of the world, are well explained by assuming that the bulk of mesons originate at the 7·5 cm. Hg pressure-level.

In addition to this we have the fact that the value of 18·6 km. for the mean range, obtained by the same hypothesis, leads to a rest life which agrees with the one generally accepted (Nereson and Rossi 1943). If we take for the mass $180\,m_e$, and for the relevant average momentum of the mesons along their path down to sea-level the value 2800 Mev/c. given by Rossi (1939) (derived from Blackett's spectrum and assuming a momentum loss of 2×10^6 ev/c. per gm/cm^2), then from $L=p\tau_0/\mu$ we have $\tau_0=2{\cdot}04\times10^{-6}$ sec.

We might therefore conclude from these results that the bulk of mesons is generated at the 7·5 cm. pressure-level.

To see, however, what the results would be if a higher pressure-level had been assumed for the layer of meson formation, all the preceding computations have been repeated taking the 3 cm. pressure-level, the highest for which sufficient meteorological data are now available. Table 4 compares the results so obtained with the previous results.

Table 4

	a_1 (%)	a_2 (%)	12-monthly var. ampl. (%)
3·0 cm. Hg pressure-level	−0·15	−0·43	2·40
7·5 cm. Hg pressure-level	−0·10	−0·24	1·65
Observed values	−0·10	−0·18	1·43

We can see that the hypothesis of meson formation at the 3 cm. pressure-level leads to values which are roughly 50% higher than those for 7·5 cm. Hg. To reconcile them with the experimental results, we should clearly have to take for the mean range of mesons a value of 28 km., a 50% increase, and this value, together with the fact that in this case we should have to take for the average momentum of the mesons along their path 2100 Mev/c. (Jánossy 1948), would lead to a rest life of $4{\cdot}0\times10^{-6}$ sec., a 100% increase over the accepted value.

The conclusion appears therefore to be that the bulk of mesons is generated at, or little above, the 7·5 cm. pressure-level, at the mean height of 16 km.

If the average layer of meson formation lies at that height, it follows that the intensity of the meson component should be decreasing at 3 cm. Hg (22 km.). On the other hand, the experimental result of Schein, Jesse and Wollan (1941) shows that the hard component increases gradually up to the height of 2 cm. Hg reached in their experiment. The two results, however, are not necessarily mutually exclusive if, as is generally supposed, the mesons are formed by primary protons the penetrating power of which is not too different from that of the hard component. The intensity of the primary protons should increase gradually with height.

In connection with the origin of the soft component, it may be significant that the layer of meson formation appears to be located at the height of the maximum of the Pfotzer curve for the total intensity (1936).

ACKNOWLEDGMENTS

The author wishes to express his indebtedness to the Director of the Meteorological Office, Air Ministry, for kindly placing the upper-air data at his disposal.

He also wishes to express gratitude to Sir George Thomson for providing facilities for this work to be carried out in the Physics Department, Imperial College of Science and Technology, and to Professor P. M. S. Blackett for his very kind interest and helpful suggestions at all times.

REFERENCES

BLACKETT, P. M. S., 1938, *Phys. Rev.*, **54**, 973.
CLAY, J., and BRUINS, E. M., 1939, *Rev. Mod. Phys.*, **11**, 158.
COMPTON, A. H., and TURNER, R. N., 1937, *Phys. Rev.*, **52**, 799.
DUPERIER, A., 1941, *Proc. Roy. Soc.* A, **177**, 204 ; 1945, *Proc. Phys. Soc.*, **57**, 464.
FORBUSH, S. E., 1938, *Phys. Rev.*, **54**, 975.
GILL, P. S., 1939, *Phys. Rev.*, **55**, 1154.
HESS, V. F., 1940, *Phys. Rev.*, **57**, 781.
HOGG, A. R., 1947, *Proc. Roy. Soc.* A, **192**, 128.
JÁNOSSY, L., 1948, *Cosmic Rays* (Oxford : University Press).
NERESON, N., and ROSSI, B., 1943, *Phys. Rev.*, **64**, 199.
PFOTZER, G., 1936, *Z. Phys.*, **102**, 23.
ROSSI, B., 1939, *Rev. Mod. Phys.*, **11**, 296.
SCHEIN, M., JESSE, W. P., and WOLLAN, E. O., 1941, *Phys. Rev.*, **59**, 615.
SCHONLAND, B. F. J., DELATIZKY, B., and GASKELL, J., 1937, *Terr. Mag.* and *Atmos. Elect.*, **42**, 137.
VALLARTA, M. S., and GODART, O., 1939, *Rev. Mod. Phys.*, **11**, 180.

The Attenuation of Ultra-High-Frequency Electromagnetic Radiation by Rocks

BY R. I. B. COOPER,
Department of Geodesy and Geophysics, University of Cambridge

Communicated by B. C. Browne ; *MS. received* 21 *November* 1947, *and in amended form* 25 *February* 1948

ABSTRACT. Radiation of frequency 200 Mc/s. was transmitted through the covering rocks of certain railway tunnels from a radar transmitter feeding a special aerial. The object was to measure the attenuation of rocks *in situ* to see if there might be practical applications to geophysical prospecting. The attenuation in dry sandstone was found to be 3 to 4 db. per foot, which was much greater than had been hoped. It thus appears that the penetration is insufficient to be of practical use.

§1. INTRODUCTION

THE use of electromagnetic radiation for geophysical prospecting, as opposed to the use of induction methods, does not appear to have been extensively discussed. A great deal of literature exists, however, on the effect of the surface conductivity of the ground on the propagation of surface waves. Recently in America attention has been focused on the possible use of surface wave measurements for detecting dykes and faults or other concealed geological discontinuities which have a considered effect on the signal strength measured near them. In Germany Grosskopf and Vogt (1940, 1942, 1943) have been principally responsible for the interpretation of effective surface conductivities in terms of a structure consisting of layers of different conductivity. In Russia, Mandelstam and

Papalexi (1938) have developed a technique which was used by Al'pert (1945) to investigate the refraction effects near a horizontal conductivity boundary.

By contrast, little attention has been drawn to the possibility of penetrating the earth by suitably directed radiation. It has usually been regarded as impossible because of the conductivity of the rocks; nevertheless, consultation of available figures suggested that some rocks might be found with sufficiently low conductivity. Jakosky (1937), dealing quantitatively with the effect of moisture content on the conductivities of different rocks, found that the conductivity depended principally on the amount of moisture present, and very little on the material of the rock. A sandstone with 2–3% water was found to have a conductivity of about 10^{-14} E.M.U.* Figures for effective surface conductivity given by Grosskopf and Vogt ranged from 9.10^{-15} to 4.10^{-13}, and they suggested that the rocks below the upper layer might have smaller conductivity. Fritsch (1943) performed some experiments on radio communication in mines at frequencies between 4 and 30 Mc/s. and concluded that the ranges achieved accorded with a smaller attenuation than was commonly expected.

The simplest way of using electromagnetic radiation for prospecting would be to measure the time required for a pulse to be transmitted to a buried object and reflected. For such a "radar" technique the frequencies used must be high enough to enable sufficiently short pulses to be generated. Setting the pulse length arbitrarily at $\frac{1}{5}$ μsec. it is seen that frequencies much below 200 Mc/s. are useless. The U.H.F. conductivity loss is nearly independent of frequency, but the dielectric losses, notably of water, increase rapidly with increasing frequency: at 200 Mc/s. these losses are still negligible in comparison with the conductivity loss, and the losses of other constituents in the main body of rock—mica, quartz, calcite, felspar, etc.—are still small compared with the loss in the water present.

A calculation given later shows that, for a conductivity of 10^{-14} and a frequency of 200 Mc/s. reflections might be obtained from within about 50 m. Losses due to scattering, imperfect reflection, aerial mismatches, etc., would affect this range, but since it is determined mainly by an attenuation it would not be reduced appreciably provided these losses did not exceed a few db. On the other hand, the range would be almost exactly inversely proportional to the conductivity. Nevertheless, in view of the lack of data and the importance of the possibilities, an experiment at a selected dry site was considered practicable.

To determine the attenuation without using reflections, a radar transmitter was mounted above a railway tunnel and provided with a special aerial for directing radiation into the ground after the removal of the surface soil. An exploring receiver was used to measure the field in the tunnel, the output being displayed on a cathode-ray tube with a high-speed trace. The tunnels investigated were: Colwall, Worcs., keuper marl; Alderton, Glos., great oolite; and Pontypridd, Glam., coal measure sandstone.

§2. EXPERIMENTAL METHOD

Essentially the method consisted in transmitting a known power downwards into the ground above the tunnel and measuring the field strength with a receiver in the tunnel after passage of the radiation through a given thickness of rock.

* E.M.U. are used throughout the paper unless otherwise stated.

The field strength E at the receiver is given approximately by

$$E(\text{v/cm.}) = (7{\cdot}0\sqrt{(PGf_Tf_R)}/r) \times e^{-kt}. \qquad \ldots\ldots(1)$$

P = power of transmitter (watts); f_T = fraction of power entering ground; f_R = fraction of power emerging from rock to reach receiver; G = aerial gain*; t = thickness of rock (cm.); r = distance from transmitter to receiver (cm.); k is defined as the attenuation constant and $1/k$ is seen to be the distance in which the field strength is reduced to $1/e$ of the value it would have in the absence of the rock ($k = 0$ *in vacuo*).

It can be shown that $2k^2 = (4\pi^2/\lambda^2)[\sqrt{(\epsilon^2 + 4\sigma^2/f^2)} - \epsilon]$, where λ = wavelength; ϵ = dielectric constant; σ = conductivity in E.S.U.; f = frequency.

At 200 Mc/s. we find, for reasonable values of ϵ and σ, that $\epsilon \gg 2\sigma/f$ and

$$1/k \simeq c\sqrt{\epsilon}/2\pi\sigma; \qquad \ldots\ldots(2)$$

with $\epsilon = 9$ and $\sigma = 10^{-14}$ E.M.U.

$$1/k \simeq 16 \text{ metres}.$$

The exponential factor in (1) is thus seen to be the governing term in determining the range of penetration to be expected and that, therefore, when solving for r, orders of magnitude only need be inserted for quantities in this expression other than k. f_T may be assumed 0·5 since we can control the method of direction into the ground, but f_R may be only of order 0·1 since for practical reasons it is impossible to provide a special receiving aerial inside a railway tunnel. Putting $E = 10^{-6}$, $PG = 10^5$ and $r = t$, we then find $r \sim 150$ metres. This distance will, however, be almost exactly inversely proportional to σ. In the event of the attenuation being low enough for reflections to be obtained, the expected maximum depth of detection would be rather less than half this figure. It is not expected that ϵ will show very wide variations and the expected range depends in any case only on its square root.

In order to determine k, the variation of E with t should have been studied. This would have involved serious practical difficulties. Either the rock would have had to be removed from the site to reduce the thickness between transmitter and receiver or else new sites would have had to be occupied with a thinner cover to the tunnel. The former alternative would have been prohibitively laborious and the latter procedure would have been invalid if the rocks were inhomogeneous in a horizontal direction; the terrain might also have proved to be unsuitable. Therefore it was resolved to measure each of the quantities in (1) separately at any site which proved suitable. It must be borne in mind that the main aim of the experiments was to discover *rapidly* whether there was any chance of the proposed technique being successful: had it so proved, more accurate measurements would have been undertaken.

Measurement of E, t and r was simple, and a theoretical figure was used for G; the measurement of P, f_T and f_R presented difficulties. It is clear that in endeavouring to direct power from a transmitter in air into the ground there will be loss due to reflection at the interface. For the purpose of these experiments the magnitude of this loss was immaterial so long as it could be measured. Therefore a special aerial was constructed (see § 3) from which power could not escape except either by entering the ground or returning down the lead to the transmitter. The power flowing both to and from the aerial was measured and the difference

* This is the ratio of the power per unit solid angle sent in the direction of the receiver to that which would be sent if the power were uniformly radiated in all directions.

therefore represented the amount entering the ground. In this way the values of P and f_T were found.

f_R was simply the transmission coefficient from the rock into the air in the tunnel since no special aerial was used. It was assumed to be the same as that at the surface where the transmitter was situated, and was deduced from measurements made on the transmitting aerial. An overall test was also made in free space by transmitting across a stretch of water and comparing the observed value of E with that calculated from P, G and r. This test gave satisfactory agreement.

§3. DETAILS OF THE EXPERIMENTAL EQUIPMENT

A block-schematic diagram of the equipment is given in figure 1.

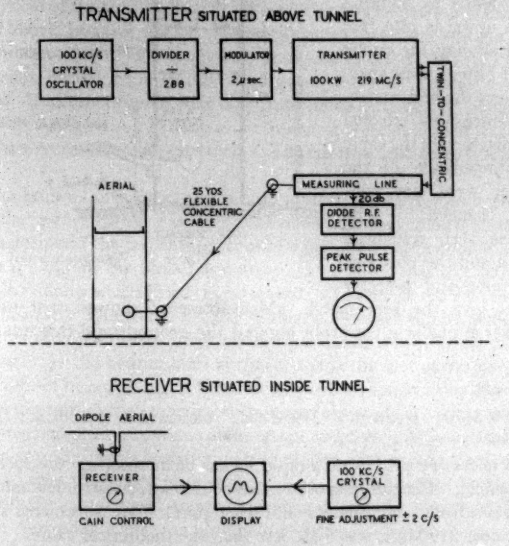


Figure 1. Block schematic diagram of apparatus.

(i) *Transmitter*

Type T 3154B; frequency 219 Mc/s.; nominal peak power 100 kw.; pulse length 2 μsec.; recurrence 3–400 c/s.

It was necessary to measure the peak power being actually transmitted into the ground and to monitor this continuously in the field. Since no wattmeter operating at this frequency was available, a concentric-line directional coupling unit was developed. This type of measuring line has only been described in Service reports; a small amount of the power passing along the line in one direction only can be coupled into a suitable detector. The transmitter output was converted from balanced twin feeder to concentric line by a suitable transformer before connection to the measuring line: from the latter, 25 yds. of flexible cable (Uniradio 34) led to the aerial. The measuring line was double-ended; it could

thus be reversed, so that the power flowing both to and from the aerial could be read. The peak power at the transmitter was usually about 75 kw.

(ii) *Transmitter aerial*

This consisted of an open-mouthed waveguide of section 45 × 15 in. about 3 ft. long and provided with a movable reflecting piston and adjustable probe, connected by concentric cable (Uniradio 34) to the end of the measuring line (figure 2). The cable attenuation was 1·8 db. The best dimensions of the probe were first found by pointing the mouth upwards into free space and altering the probe length and piston-probe spacing to achieve the best match. As the dimensions affected only the matching of the cable into the impedance of the guide, they were not subsequently changed. The mouth was then rested on the ground and the positions of both probe and reflector adjusted to give the best match. Measurements showed that usually 75% of the power available at the probe entered the ground, and this was considered adequate.

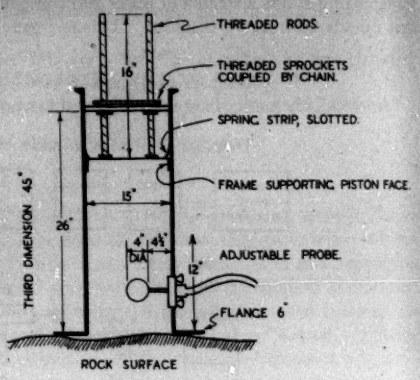


Figure 2. Waveguide aerial.

(iii) *Receiver*

Type R 3170; frequency 219 Mc/s.; bandwidth 1·4 Mc/s.; noise level $0{\cdot}5 \times 10^{-12}$ w.

It was necessary to provide a rapid means of monitoring the receiver noise-level in the field. For this purpose a noise-source was constructed using a CV 172 diode and run from accumulators (B.T.H. report). Measurements showed that its output near 219 Mc/s. was 3 db. less than the theoretical value.

The receiver gain was calibrated by well known methods (Cooper and Freeman 1943, Varley *et al.* 1943).

(iv) *Receiver aerial and display*

The receiver aerial was a simple dipole mounted on a bakelite rod and fed from a few yards of flexible cable so that it could be used to explore the field. Its matching into the receiver was determined with the aid of the measuring line. It was found that at most 10% of the power was reflected.

In order to display the received pulses without introducing a connection between transmitter and receiver down which radiation might leak, and at the same time to have a time-base of sufficiently high speed to enable small time delays to be detected, it was necessary to control the recurrence of both transmitter and receiver from crystals. The transmitter recurrence was derived from a 100-kc/s. crystal by repeated division with multivibrators, and the receiver time-base

consisted of a triangular waveform with a forward strioke of 7 μsec. directly generated from another 100-kc/s. crystal. A small trimmer enabled the display to be synchronized with the transmitter to within $\frac{1}{10}$ c/s. In this way the time-distribution of events could be studied sufficiently exactly to distinguish the direct signal from that leaking round the tunnel mouth.

§4. RESULTS

The selection of the tunnels was a matter of some difficulty. The main consideration was that they should be dry, but they also had to have convenient access by road, both for the transmitter in its vehicle above the tunnel and for the receiver and its generator, which had to be run in on a trolley from the nearest station. Traffic density was also a factor in the selection.

The dryness of the tunnels could only be judged by inspection of the walls and the ground above, but this did not necessarily mean that water was not present in the intervening rock. The experiments were conducted in August 1947, which was the driest period experienced for several decades. However, neither in Colwall nor Alderton tunnels were any indications obtained of signals arriving other than by leakage round the mouth: it can only be supposed that the intervening rocks were wetter than expected. The cover in both cases was about 40 ft. thick.

In Colwall tunnel sites were tried at both western and eastern ends; the western end was rather wet (a spring was later discovered near the tunnel mouth), whilst at the drier eastern end the transmitter could be mounted only on the road immediately above the line of the tunnel and about 35 yds. back from the mouth, where considerable radiation leakage was experienced. This made the field in the tunnel very complicated. The traffic-density in this single-track tunnel made thorough exploration of the field difficult, but no obvious indications of direct transmission were noticed. As the keuper marl at this end is at the foot of the steep Malvern Hills it would not be surprising if it were damper than superficially appeared.

In Alderton tunnel the cover is constant and about 40 ft. thick, and consists of very inhomogeneous intercalations of limestone and clay. Although this appeared very dry, it is quite possible that much scattering and multiple reflection could occur. Two sites were tried immediately above the line of the tunnel, and about 100 yds. from the eastern end, to see if accidental imperfections were affecting the penetration, but no success was obtained at either.

It is clear that the efficiency of transfer of the radiation into the air in the tunnel from the rocks is low. Not only is the transmission coefficient small, but the curved tunnel roof acts as a lens tending to spread the beam and reduce the field strength. From measurements on the transmitting aerial a transmission coefficient of 0·2 was obtained, and from this the refractive index of the ground could be estimated: it was thus possible to calculate the "lens effect" from the dimensions of the tunnel. In the case of Alderton and Pontypridd tunnels the total calculated loss was about 25 db.

At Pontypridd the site was considerably better: the cover increased gradually so that at the site 100 yds. inside the northern mouth it was only 33 ft. thick; it consisted of massive homogeneous dry sandstone and the tunnel was almost disused, so that it was possible to run the receiver up and down the line on a trolley and observe and plot the signal strength without waiting for the passage of trains and clearance of smoke.

In figure 3 is shown the variation with distance into the tunnel of the signal strength, assessed as the height of the largest pulse visible. It is believed that the full curve shows the decay of the signals leaking into the mouth, and the dotted curve that of signals received directly through the roof. It will be seen that the observations show a greater scatter in the region below the transmitter and

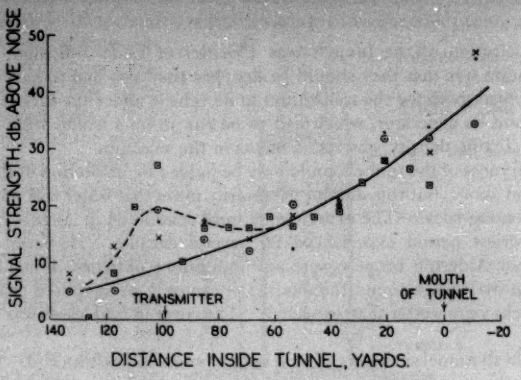


Figure 3. Variation of signal strength in Pontypridd tunnel.
• 1st run, Sept. 9. Polarized ∥ rails. × 3rd run Sept. 9. Polarized ∥ rails.
○ 2nd run „ „ „ ⊥ „ ⊡ 4th run „ 10. „ ∥ „

near the mouth: this may be because here there are strong interference fields, and small movements of the aerial, such as might easily occur in between runs, can cause large changes in signal strength. Nevertheless it is considered that the small maximum below the transmitter is significant (the distance of the transmitter from the mouth was not known until after the measurements had been completed). The aerial was mounted about 3 ft. above the offside rail: in the non-interfering zone the strength did not depend in a regular manner on position or orientation provided the aerial was not too close to floor or walls. Runs were made with the aerial polarized horizontally both perpendicular and parallel to the rails, but no outstanding difference was observable.

The trace was also observed whilst the trolley was run back and forth under the transmitter site: a particular echo on the trace was observed to grow and decay during this test, which was very convincing. The appearance of the trace is sketched in figure 4. A small pulse is seen before the direct signal, which should in this case be the earliest arrival, but since the time-base runs continuously this may in fact be an echo arriving on the next trace, 10 μsec. later. This would be possible since many large reflecting objects were situated at the required range from the tunnel mouth. Ignoring this pulse, therefore, the next pulse seen in figure 4 is delayed by approximately 1 μsec. on the direct pulse,

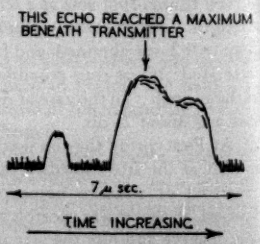


Figure 4. Appearance of trace with signals.

or about 350 yds.—a distance of the right order for a path round the tunnel mouth.

The signals were strongly modulated by the approach of trains.

§5. CALCULATION OF ATTENUATION

The transmitter power is $7{\cdot}5 \cdot 10^4$ w. and aerial gain ~ 5; the power flux at 1 yd. distance is therefore $7{\cdot}5 \cdot 10^4 (5/4\pi)(91{\cdot}44)^2 = 3{\cdot}56$ w/cm^2. This is equivalent to a field strength of 69·4 v/cm.

The observed signal in the tunnel at 16 yds. range was 10 db. above noise. The noise-level of the receiver at the time of the experiment was $4 \cdot 10^{-13}$ w.

A field of strength f v/cm. produces power $\lambda^2 f^2/320\pi^2$ w. in a receiver matched to a simple dipole. Hence in this case, since $\lambda = 135$ cm., we find $f = 0{\cdot}838 \cdot 10^{-6}$ v/cm. The field expected in free space at 16 yds. range is $69{\cdot}4/16 = 4{\cdot}34$ v/cm. Hence the attenuation to be accounted for is $4{\cdot}34/0{\cdot}838 \cdot 10^{-6}$ or 134·3 db. Allowing 3 db. loss at the transmitter and 25 db. for the effects where the wave enters the tunnel, we find the attenuation of Pennant sandstone to be 3·2 db/ft. This is about 20 times greater than that expected assuming $\sigma = 10^{-14}$ E.M.U., so that the conductivity must have been about 2×10^{-13} E.M.U. (resistivity 5000 Ω/cm.). According to Jakosky this could have been accounted for by the presence of 15% of water, assuming Pennant sandstone to be similar to his reservoir sandstone. It seems therefore that even after prolonged surface dryness the rocks near the surface may be quite damp.

§6. CONCLUSION

The penetration of ultra-high-frequency electromagnetic radiation into rocks is insufficient to afford a practical method of geophysical prospecting. This is probably because of the great power of rocks to occlude and retain moisture even after prolonged drought. In cases where there was appreciable scattering due to inhomogeneous ground there would be even less chance of success.

Increase of power and gain would not increase the range greatly since the attenuation is exponential. Sufficient penetration could not be obtained by reduction of frequency since pulse widths would then be too great for the observation of reflections.

ACKNOWLEDGMENTS

The success of the experiments was largely due to the full co-operation of the G.W.R., and the author wishes to express here his sincere gratitude to them. The radar set was lent by kind permission of the Radar Research and Development Establishment, Ministry of Supply, Malvern, who gave laboratory facilities for the calibrations, and to whom also the author is very much indebted.

REFERENCES

AL'PERT, J. L., and GOROZHANKIN, H., 1945, *J. Phys., U.S.S.R.*, **9**, 115.
BIRCH, F., 1942, *Handbook of Physical Constants*, Geol. Soc. Amer. Spec. Paper No. 36.
B.T.H., Special Report No. L293S.
COOPER, R. I. B., and FREEMAN, J. Y., R.R.D.E., Ministry of Supply, Research Report 251.
FRITSCH, V., 1943, *Hochfrequenztech. u. Elecktroakust.*, **62**, 50.
GROSSKOPF, J., 1942, *Hochfrequenztech. u. Elektroakust.*, **60**, 136.
GROSSKOPF, J., and VOGT, K., 1940, *Telegr.- u. Fernspr. Tech.*, **29**, 164; 1943, *Hochfrequenztech. u. Elektroakust.*, **62**, 14.
JAKOSKY, J. J., 1937, *Geophysics*, **2**, 33.
MANDELSTAM, L., 1938, *Izv. Akad. Nauk. S.S.S.R.*, 525.
MEAD, L., R.R.D.E., Ministry of Supply, Research Report 268.
PAPALEXI, N., 1938, *Izv. Akad. Nauk. S.S.S.R.*, 539.
VARLEY, G. C., A.O.R.G. Memorandum No. 467.

The Application of a Variational Method to the Calculation of Radio Wave Propagation Curves for an Arbitrary Refractive Index Profile in the Atmosphere

By G. G. MACFARLANE

Telecommunications Research Establishment, Ministry of Supply, Gt. Malvern

MS. received 23 *October* 1947

ABSTRACT. The problem of determining the field strength produced by a microwave transmitter beyond the horizon has been intensively studied in recent years. Charts and tables, which are based on approximate solutions of the wave equation obtained by the phase-integral method, have been produced (Booker and Walkinshaw 1946), from which the field strength can be calculated when the modified refractive index of the air (the M-curve) varies with height according to a power law. In many cases, however, the M-curve cannot be adequately represented by a power law, and serious analytical difficulties are encountered if the problem is tackled by the phase-integral method.

The basic problem is to find the eigenvalues D_m and eigenfunctions U_m of the wave equation $d^2U_m/ds^2+\{s+f(s)+D_m\}U_m=0$, in which $f(s)$, the M-anomaly, is a given function, which tends to zero as s tends to infinity. In this paper a method, previously described by the author (Macfarlane 1947), is applied to determine the first eigenvalue D_1 for the class of M-anomalies that can be represented by a series of the form, $\sum_{n=1}^{r} A_n \exp(-\alpha_n s)$. Two numerical examples are given to illustrate the method. They were obtained by fitting one exponential and three exponentials in turn to an M-anomaly calculated from meteorological observations made at Kaikoura, New Zealand. In the first case the M-anomaly is represented by a surface duct and in the second case by an elevated duct. The exponential attenuation rate, which is determined by the imaginary part of D_1, is calculated in both cases as a function of the wavelength.

Finally, the theoretical height–gain function for a wavelength of 3 metres is calculated from the M-curve observed 47 miles off shore from Ashburton, South Island, New Zealand, in the late afternoon of 4th November 1946. It is found that there is close agreement between theoretical and observed height–gain curves.

§1. INTRODUCTION

THE theory of propagation in current use for describing the field beyond the horizon of a radio transmitter expresses the field as a sum of normal modes of the wave equation. The independent variables of the wave equation, which are height and range, are separable, since the atmosphere is assumed to be horizontally stratified. This simplification reduces the problem of solving the wave equation with two independent variables to that of finding the eigenvalues and normal modes of the one-dimensional wave equation *

$$d^2U_m/dh^2+k^2\{y(h)+\Lambda_m\}U_m=0. \qquad \ldots\ldots(1)$$

Each mode must satisfy the boundary conditions, that $U_m(0)=0$, and that as $h\to\infty$ U_m represents an upgoing wave. In this equation

$$y(h)=N^2(h)-1=2\times10^{-6}M(h), \qquad \ldots\ldots(2)$$

where $N(h)$ is the modified refractive index of the air and Λ_m is the mth eigenvalue.

* Since the subject matter of this paper is closely related to that of Pekeris (1946) we shall adopt a similar notation.

A number of different methods have been used to evaluate U_m and Λ_m. Booker and Walkinshaw (1946) applied the phase-integral method, originally discussed in this connection by Eckersley and Millington (1938), to find U_m and Λ_m for the M-curve,

$$M(h) = K_\infty(h - d^{1-m}h^m/m) \times 10^{-6}, \qquad \ldots\ldots(3)$$

where $m = 0{\cdot}5$. Here d is the duct width and K_∞ is the gradient of modified refractive index at infinite height.

The phase-integral method leads to solutions which are only approximately correct but which, nevertheless, may be sufficiently accurate for most practical purposes; the approximation that the atmosphere is a horizontally stratified medium is indeed less satisfactory than the analytical approximation made in the phase-integral theory. However, the phase-integral method leads to troublesome mathematical problems when the M-curve has a form more complicated than that of equation (3). Even when $m = 0{\cdot}2$ it was too complicated to evaluate analytically and it was found necessary to compute solutions by means of a differential analyser. Techniques for evaluating the required solutions of the wave equation on the differential analyser were devised by Hartree, Michel and Nicolson (1946).

Another method of evaluating U_m and Λ_m has been described in a recent paper by Pekeris (1946). The M-curve is expressed in the form

$$M(h) = K_\infty h + g\epsilon^{-ch}$$

and U_m and Λ_m are found by a perturbation method. The procedure is to express U_m as a linear combination of the eigenfunctions U_n^0 of all the modes in the standard case, in which $M(h) = K_\infty h$. The coefficients A_{mn} of U_n^0 and the eigenvalues Λ_m are found as solutions of an infinite system of equations. They are evaluated by an iterative procedure, which has been found to be rapidly convergent. Pekeris has also shown that a more general class of M-curves can be handled by a slight extension of the perturbation method. It includes M-curves of the forms

$$M(h) = K_\infty h + \sum_n g_n \exp(-c_n h) \qquad \ldots\ldots(4.1)$$

and

$$M(h) = K_\infty h + \sum_n g_n h^n \exp(-c_n h). \qquad \ldots\ldots(4.2)$$

Since an arbitrary M-curve can be represented by either of those series to any desired degree of accuracy, it might be said that the general problem has been solved. On the other hand the convergence of the procedure is least rapid when the mode is slightly leaky and this is a case of great interest.

The procedure we propose is based on a variational method in which the eigenvalue is found from the solution of a transcendental equation. It is not easy to estimate the accuracy of the eigenvalues obtained, but the eigenvalues found by it for the power-law M-curve of equation (3) agreed within 0·4% with those obtained by Hartree, Michel and Nicolson on the differential analyser (Macfarlane 1947).

We have developed the method for an M-curve of the form

$$M(h) = K_\infty h + g_0 + \sum_{n=1}^{r} g_n \exp(-c_n h), \qquad \ldots\ldots(4.3)$$

so that it can be applied to any M-curve. If it is desired, an M-curve expressed in the form of equation (4.2) can also be handled with the information given in this paper.

Following Pekeris (1946) we introduce natural units of height

$$\left.\begin{aligned} s &= h/H \qquad H = (k^2q)^{-\frac{1}{3}} \\ q &= dN^2/dh = 2K_\infty \times 10^{-6} = 2{\cdot}5 \times 10^{-9}\ \text{cm}^{-1}\ * \\ D_m &= \Lambda_m (k/q)^{\frac{1}{3}}, \end{aligned}\right\} \qquad \ldots\ldots(5)$$

so that equation (1) assumes the form

$$d^2U_m/ds^2 + \{s + f(s) + D_m\}U_m = 0. \qquad \ldots\ldots(6)$$

In terms of s the M-curve, equation (4), gives

$$f(s) = A_0 + \sum_{n=1}^{r} A_n \exp(-\alpha_n s), \qquad \ldots\ldots(7)$$

where

$$A_n = wg_n, \qquad \ldots\ldots(7.1)$$

$$\left.\begin{aligned} \alpha_n &= 2 \times 10^{-6} c_n/wq, \\ &= 26{\cdot}25\, c_n/w, \quad c_n \text{ in } ft^{-1}, \end{aligned}\right\} \qquad \ldots\ldots(7.2)$$

and

$$\begin{aligned} w &= 2 \times 10^{-6} (k/q)^{\frac{1}{3}} \qquad \ldots\ldots(7.3) \\ &= 3{\cdot}696 \lambda^{-\frac{1}{3}}, \quad \lambda \text{ in cm}. \end{aligned}$$

In the work to be reported here we have confined our attention to the determination of the first eigenvalue D_1 and its eigenfunction $U_1(s)$. This has been done because, in many cases of importance, the contributions of higher modes to the field strength beyond the horizon is negligible. In terms of natural units of horizontal distance d,

$$x = d/L, \quad L = 2(kq^2)^{-\frac{1}{3}}, \qquad \ldots\ldots(8)$$

and the field strength at range x and height s_2 due to a transmitter of power gain G and power P at height s_1 is

$$E = \frac{(120\pi PG)^{\frac{1}{2}}}{L} x^{-\frac{1}{2}} \sum_m \exp(iD_m x) U_m(s_1) U_m(s_2) \qquad \ldots\ldots(9.1)$$

$$\simeq \frac{(120\pi PG)^{\frac{1}{2}}}{L} x^{-\frac{1}{2}} \exp(iD_1 x) U_1(s_1) U_1(s_2) \qquad \ldots\ldots(9.2)$$

at sufficiently great range beyond the horizon. In these formulae the eigenfunction U_m is normalized so that

$$\int_0^\infty U_m^2(s)ds = 1. \qquad \ldots\ldots(10)$$

From equations (7.3) and (8) the numerical distance is expressible in terms of the wavelength parameter w as

$$x = 0{\cdot}1423\, w^{\frac{3}{2}} d, \quad d \text{ in miles}. \qquad \ldots\ldots(11)$$

The exponential attenuation rate of the first mode is, therefore,

$$8{\cdot}686\, xd^{-1} \text{Im}(D_1) = 1{\cdot}236\, w^{\frac{3}{2}} \text{Im}(D_1) \text{ db. per mile}. \qquad \ldots\ldots(12)$$

§ 2. THE VARIATIONAL METHOD

The variational equations of the wave equation, equation (6), in which the function $f(s)$ is given by equation (7), are (Macfarlane 1947)

$$D(\phi) = \int_0^{\infty\epsilon} \left(\frac{d\phi}{dy}\right)^2 dy - \int_0^{\infty\epsilon} y\phi^2 dy - \sum_{n=1}^{r} A_n \int_0^{\infty\epsilon} \exp(-\alpha_n y)\phi^2 dy \qquad \ldots\ldots(13.1)$$

and

$$H(\phi) = \int_0^{\infty\epsilon} \phi^2 dy. \qquad \ldots\ldots(13.2)$$

* We take the "effective" radius of the earth to be 8×10^6 metres.

It should be noticed that the constant term A_0 is absorbed into D_1. D_1+A_0 is then the minimum stationary value of the ratio D/H for functions ϕ that satisfy the boundary conditions (i) $\phi(0)=0$; (ii) $\phi(y)\to 0$ as $y\to\infty\epsilon$, where $-\pi/3<\arg\epsilon<0$, and that have no zero values on the contour from $y=0$ to $y=\infty\epsilon$. As explained by Eckersley and Millington (1938) a suitable choice for ϕ is the Airy function

$$\phi=\mathrm{Ai}(cy-j_1), \qquad \ldots\ldots(14)$$

where $-j_1=-2{\cdot}3381$ is the first zero.

Parameter c is determined so that the ratio D/H has a minimum stationary value and $\epsilon=1/c$.

Substitution of function ϕ, given by equation (14), into the variational equations (13) gives

$$\frac{D(\phi)}{H(\phi)}=\frac{1}{Q}\left[P\left(c^3-\frac{2}{c}\right)-\sum_{n=1}^{r}A_nF(\alpha_n/c)\right], \qquad \ldots\ldots(15)$$

where

$$P=\int_0^\infty\{\mathrm{Ai}'(y-j_1)\}^2dy=0{\cdot}38321\,;\quad Q=\int_0^\infty\{\mathrm{Ai}\,(y-j_1)\}^2dy=0{\cdot}49170 \qquad \ldots\ldots(16)$$

$$F(z)=\int_0^\infty\exp(-zy)\{\mathrm{Ai}\,(y-j_1)\}^2dy.$$

The stationary values of the ratio D/H occur at the saddle points of the function D/H in the complex c-plane; that is, where

$$\frac{d}{dc}\left(\frac{D}{H}\right)=\frac{1}{Q}\left[2P\left(c+\frac{1}{c^2}\right)+\frac{1}{c^3}\sum_{n=1}^{r}A_n\alpha_nF'(\alpha_n/c)\right]=0$$

or

$$0=c^3+1+1{\cdot}3048\sum_{n=1}^{r}A_n\alpha_nF'(\alpha_n/c). \qquad \ldots\ldots(17)$$

The root of equation (17) lying in the fourth quadrant which gives the smallest value for the modulus of D/H, is the required value of c. On substitution of this value of c into equation (15) the approximate value of D_1+A_0 is obtained. It should be noted that the constant term A_0 affects only the phase of the first mode, since, by equation (12), the attenuation rate is dependent solely on the imaginary part of D_1.

The application of the variational method to the calculation of D_1 for an arbitrary M-curve involves therefore four steps:

(i) representation of the given M-curve by an expression of the form

$$M(h)=K_\infty h+g_0+\sum_{n=1}^{r}g_n\exp(-c_nh) \qquad \ldots\ldots(18)$$

with r, the number of exponential terms, as small as possible;

(ii) calculation of the Laplace transform of the square of the Airy function and of its derivative for complex values of the argument z,

$$F(z)=\int_0^\infty\exp(-zy)\{\mathrm{Ai}(y-j_1)\}^2dy \qquad \ldots\ldots(19)$$

and

$$F'(z)=-\int_0^\infty y\exp(-zy)\{\mathrm{Ai}(y-j_1)\}^2dy; \qquad \ldots\ldots(20)$$

(iii) solution of the minimal variation equation (17) to find c;

(iv) calculation of $A_0+D_1=D/H$ from equation (15) using the values of c found from (iii).

When D_1 has been found the corresponding eigenfunction $U_1(s)$ is obtained by numerical integration or by use of a differential analyser.

For (i) one method of fitting curves of the form of equation (18) to given M-curves depends on putting $c_n = nc$ and rewriting equation (18) in the form

$$M(h) - K_\infty h - g_0 = \sum_{n=1}^{r} g_n \exp(-nch) = \sum_{n=1}^{r} g_n z^n, \text{ where } z = \exp(-ch).$$

The range of the variable z extends from 0 to 1. Since K_∞ and g_0 can be obtained from the asymptotic behaviour of $M(h)$ for large values of h, the problem becomes one of fitting a polynomial to a given set of points, in the range $0 \leqslant z \leqslant 1$. Gram polynomials, which have been extensively tabulated by Davis (1935), are admirably suited to this purpose.

For (ii) the functions $F(z)$ and $F'(z)$ have been calculated to five figures at intervals of 0·2 in the real and imaginary parts of z in the range $x = 0{\cdot}0(0{\cdot}2)4{\cdot}0$, $y = 0{\cdot}0(0{\cdot}2)3{\cdot}2$, where $z = x + iy$. Both real and imaginary parts are given in a paper by Hay (1947) together with coupled differences and asymptotic formulae for extending the range of the variables.

§ 3. NUMERICAL EXAMPLES

In order to illustrate the technique we shall consider the case of an M-curve calculated from meteorological observations made at 1941 hours on 10 October

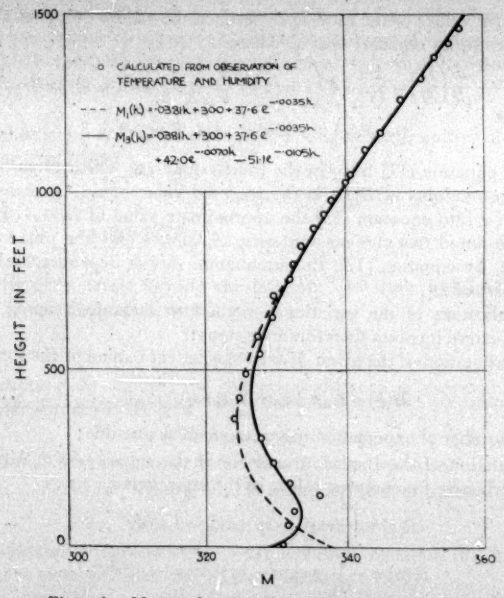


Figure 1. M-curve found at 1941 hrs. on 10 October 1944 at Kaikoura, New Zealand.

1944 at Kaikoura, New Zealand. This profile of modified refractive index can be represented by the following cubic function of exponentials *

$$M(h) = 0{\cdot}0381h + 300 + 37{\cdot}4e^{-0{\cdot}0035h} + 42{\cdot}0e^{-0{\cdot}0070h} - 51{\cdot}1e^{-0{\cdot}0105h} \quad \ldots\ldots(21)$$

* The author is indebted to Mr. J. W. Head for this formula.

The original M-curve and the approximation represented by equation (19) are shown in figure 1. In the same figure the approximation using only the first exponential term in equation (21) is also shown. In figure 2 the M-excess, denoted by μ and equal to $M(h)-0{\cdot}0381h-300$, is plotted together with the approximations represented by one and three exponential terms. This shows more clearly than figure 1 the degree of fit achieved. An even closer fit using five exponential terms is also shown in figure 2, but we shall not make use of it.

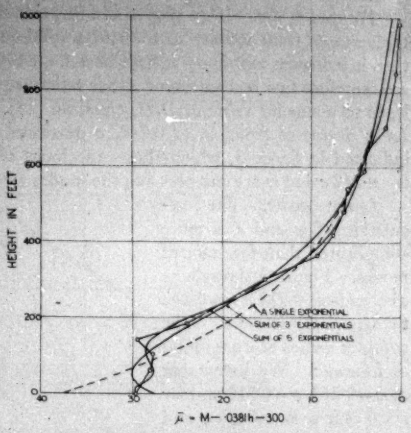


Figure 2. M-excess for M-curve of figure 1.

Figure 1 shows how one exponential term can represent a surface duct and how the addition of other two exponential terms can give rise to an elevated duct. As it is interesting to compare the eigenvalues for these two cases we shall illustrate our technique by working out both.

The first step is to find the constants A_n and α_n of the series for $f(s)$. These are given in terms of the wavelength parameter w by equations (7.1) and (7.2). Thus

$$\left.\begin{array}{llll} A_0 = & 300w & & \\ A_1 = & 37{\cdot}4w & \alpha_1 = 0{\cdot}0918/w & A_1\alpha_1 = \quad 3{\cdot}433 \\ A_2 = & 42{\cdot}0w & \alpha_2 = 0{\cdot}1836/w & A_2\alpha_2 = \quad 7{\cdot}711 \\ A_3 = & -51{\cdot}1w & \alpha_3 = 0{\cdot}2754/w & A_3\alpha_3 = -14{\cdot}073. \end{array}\right\} \quad \ldots\ldots(22)$$

We now consider separately the cases of the surface duct and the elevated duct.

(i) *The surface duct*

For the surface duct

$$M(h) = 0{\cdot}0381h + 300 + 37{\cdot}4e^{-0{\cdot}0035h} \quad \ldots\ldots(23)$$

and A_1 and α_1 have the values given in equation (22).

Equation (17), which determines c, is

$$c^3 + 1 + 4{\cdot}479F'(0{\cdot}0918/cw) = 0. \quad \ldots\ldots(24)$$

To solve equation (24) put

$$z = 0{\cdot}0918/cw \quad \ldots\ldots(25)$$

and rewrite it as

$$(0{\cdot}0918/w)^3 = -z^3\{1 + 4{\cdot}479F'(z)\}. \quad \ldots\ldots(26)$$

Since w is entirely real the problem is to find values of z which make the imaginary part of the right-hand side of equation (26) zero. Moreover, since $\arg z = -\arg c$ and since only values of c are allowed for which $0 < \arg c < \pi/3$ we need consider only values of z for which $-\pi/3 < \arg z < 0$. It is best to consider trapped modes, for which w, c and z are all real, separately from leaky modes, for which c and z are complex.

Trapped modes.—If we plot w as a function of z, a real variable, we find that w decreases from infinity at $z=0$ to a minimum value of 0·242 at $z=0·518$; then it increases to infinity again where $1+4·479F'(z)=0$. For $w>0·242$ there are therefore two real values of z that satisfy equation (24), but the smaller one leads to a smaller value of D/H, equation (15), and is the correct one to take. As w decreases from ∞ to 0·242, c decreases from 1·345 to 0·732. This is indicated in figure 3, which shows the locus of c in the complex plane. For $w<0·242$ c and w are complex and the modes are leaky.

Leaky modes. The calculation of w and c is more complicated than for trapped modes. Figure 3 shows how c varies from 0·732 at $w=0·242$ to exp $(i\pi/3)$ at $w=0$. The complex values of c are found as follows: We know that at $w=0·242$, $z=0·518$ and for $w<0·242$, z is complex and $-\pi/3<\arg z<0$. Therefore, calculate the real and imaginary parts of $-z^3\{1+4·479F'(z)\}$ at the four points (0·4, −0·2) (0·5, −0·2) (0·6, −0·2) (0·7, −0·2). These points are equally spaced, they lie along a line parallel to the real axis, and they enclose the value 0·518 between their real parts. Inverse interpolation, as described by Davis (1935), is used to find the real part of z, x_0 say, at which $\mathrm{Im}[-z^3\{1+4·479F'(z)\}]$ is zero. The values of $\mathrm{Re}[-z^3\{1+4·479F'(z)\}]$ are then interpolated to find the value at the point $(x_0, -0·2)$, which, from equation (26), is equal to $(0·0918/w)^3$. Thus w is evaluated and c is calculated from equation (25), viz. $c=0·0918/wz$. Table 1 gives details of the calculation.

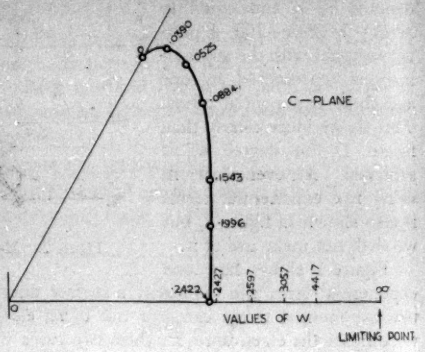


Figure 3. Locus of parameter c in complex plane.

Table 1. Evaluation of w and c for $z=x_0-0·2i$

z		$F'(z)$		$-(1+4·479F'(z))$		$-z^3(1+4·479F'(z))$	
x	y	Re	Im	Re	Im	Re	Im
0·4	−0·2	−0·34328	−0·12298	0·5823	0·5508	0·0578	−0·0424
0·5	−0·2	−0·30043	−0·10163	0·3456	0·4552	0·0871	−0·0195
0·6	−0·2	−0·25654	−0·08437	0·1490	0·3779	0·1001	0·0234
0·7	−0·2	−0·21993	−0·07036	−0·0149	0·3151	0·0863	0·0859
0·5512	−0·2					0·0956	0·0000

Therefore $w=0·2008$, $c=0·7329+0·2659i$.

This procedure is repeated for $z=x-0·4i$, $x-0·6i$ and so on until the entirely locus of c in the complex plane has been found. D_1 is evaluated for each trio of values w, c and z by substitution into equation (15), which becomes

$$D_1+300w=D/H=0·7793(c^2-2/c)-76·06F(z). \quad \ldots\ldots(27)$$

Next the exponential attenuation rate $1{\cdot}236w^{\frac{1}{2}}\mathrm{Im}(D_1)$ is calculated for each value of w. Finally the wavelength corresponding to each value of w is read off figure 4, which is a plot of the function $3{\cdot}696\lambda^{-\frac{1}{2}}$.

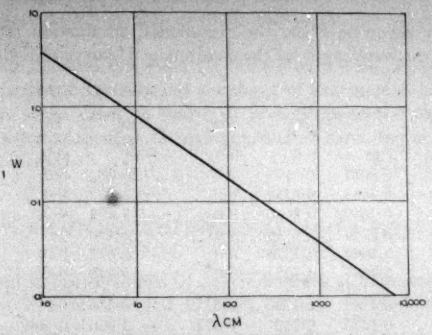


Figure 4. Parameter w as function of wavelength.

Table 2 gives the final results.

Table 2. The eigenvalues and the exponential rates as functions of the wavelength for the dotted M-curve of figure 1

z		w	c		$300w+D_1=D/H$		γ	λ (cm.)
Re	Im		Re	Im	Re	Im		
0·0	0·0	∞	1·335	0·0	$-\infty$	0·0	0·0	0·0
0·2	0·0	0·4117	1·115	0·0	−11·87	0·0	0·0	26·5
0·3	0·0	0·3057	1·001	0·0	− 8·13	0·0	0·0	42
0·4	0·0	0·2597	0·884	0·0	− 6·59	0·0	0·0	54
0·518	0·0	0·2422	0·732	0·0	− 6·04	0·0	0·0	59·6
0·5512	−0·2	0·2008	0·7329	0·2659	− 4·795	0·100	0·055	79
0·6280	−0·4	0·1412	0·7365	0·4691	− 3·094	0·456	0·212	134
0·7180	−0·6	0·1034	0·7284	0·6087	− 2·111	0·835	0·332	213
0·8067	−0·8	0·0803	0·7144	0·7085	− 1·594	1·148	0·402	311
0·9680	−1·2	0·0561	0·6664	0·8262	− 1·198	1·556	0·456	533
1·1139	−1·6	0·0438	0·6177	0·8758	− 1·178	1·775	0·459	775
1·262	−2·0	0·0364	0·5694	0·9026	− 1·230	1·896	0·447	1022
∞	$-\infty$	0·0	0·5000	0·8660	− 1·169	2·025	0·0	∞

γ=exponential attenuation rate $1{\cdot}236w^{\frac{1}{2}}\,\mathrm{Im}(D_1)$

We shall now consider the better approximation to the experimental values of M that is given by three exponential terms.

(ii) *The elevated duct*

For the elevated duct

$$M(h)=0{\cdot}0381h+300+37{\cdot}4e^{-0{\cdot}0035h}+42{\cdot}0e^{-0{\cdot}0070h}-51{\cdot}1e^{-0{\cdot}0105h}. \quad \ldots\ldots(28)$$

In place of equation (26) we now get

$$(0{\cdot}0918/w)^3=-z^3\{1+4{\cdot}479F'(z)+10{\cdot}061F'(2z)-18{\cdot}362F'(3z)\}. \quad \ldots\ldots(29)$$

As before, we have to find values of z that make the imaginary part of the right-hand side of equation (29) zero. The procedure is the same as in (i) above and the final result is given in table 3.

Table 3. The eigenvalues and the exponential attenuation rates as functions of the wavelength for the continuous M-curve of figure 1

z		w	c		$300w+D_1=D/H$		γ	λ (cm.)
Re	Im		Re	Im	Re	Im		
0·177	0·0	∞	0·0	0·0	$-\infty$	0·0	0·0	0·0
0·2	0·0	0·8085	0·5677	0·0	−26·28	0·0	0·0	9·8
0·3	0·0	0·3663	0·8354	0·0	−11·20	0·0	0·0	32·0
0·4	0·0	0·2669	0·8600	0·0	− 7·692	0·0	0·0	51·5
0·5	0·0	0·2254	0·8146	0·0	− 6·490	0·0	0·0	66·4
0·6	0·0	0·2089	0·7324	0·0	− 5·672	0·0	0·0	74·4
0·625	0·0	0·2084	0·705	0·0	− 5·648	0·0	0·0	74·7
0·6586	−0·2	0·1748	0·7300	0·2217	− 4·558	0·075	0·039	97·2
0·723	−0·4	0·1275	0·762	0·422	− 3·036	0·369	0·163	156
0·7956	−0·6	0·0945	0·779	0·587	− 2·040	0·751	0·285	244
0·867	−0·8	0·0749	0·764	0·705	− 1·479	1·081	0·366	347
0·930	−1·0	0·0626	0·732	0·786	− 1·280	1·332	0·412	440
1·107	−1·6	0·0430	0·624	0·902	− 1·178	1·742	0·447	790
∞	$-\infty$	0·0	0·500	0·866	− 1·169	2·025	0·0	∞

γ=exponential attenuation rate $1{\cdot}236w$ Im (D_1).

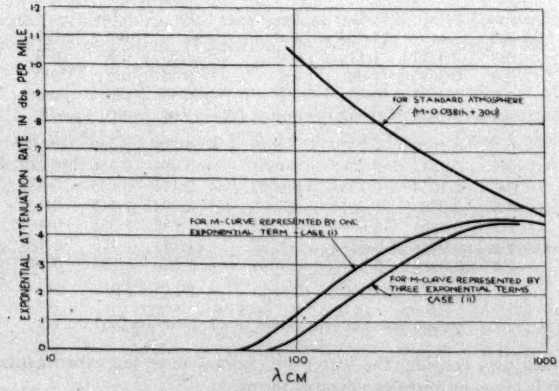

EXPONENTIAL ATTENUATION RATE IN dbs PER MILE
FOR STANDARD ATMOSPHERE
(M=0·0318h + 30h)
FOR M-CURVE REPRESENTED BY ONE EXPONENTIAL TERM - CASE (i)
FOR M-CURVE REPRESENTED BY THREE EXPONENTIAL TERMS CASE (ii)
10
100
1000
λ CM

Figure 5. Dependence of exponential attenuation rate on wavelength.

Finally, in figure 5 we have shown the dependence of exponential attenuation rate on wavelength for the approximations to the experimental M-curve of figure 1 given by the surface and the elevated ducts.

The curve representing a standard atmosphere is also shown. It is seen that the attenuation rate is greater in case (i) than in case (ii). For $\lambda > 10$ metres

there is very little difference between all three curves but for $\lambda<6$ metres the attenuation rate for the standard atmosphere is much higher than for the atmosphere represented by the curves of figure 1. In considering the curves of figure 5 it should not be forgotten that the total attenuation includes that due to the factor $x^{-\frac{1}{2}}$ of equation (9.2) as well as that due to the exponential attenuation rate, which is shown.

It must be emphasized that cases (i) and (ii) have been considered in detail primarily to illustrate the power and simplicity of this method for determining eigenvalues. In some cases as many as seven exponential terms might be required to represent adequately the experimental M-curve, but even so the extra labour involved is small.

§4. COMPARISON OF THEORETICAL AND EXPERIMENTAL HEIGHT-GAIN CURVES

The observed M-curve, shown by dots in figure 6, and the height–gain curve, shown in figure 7, were observed 47 miles off shore from Ashburton, South Island, New Zealand, in the late afternoon of 4 November 1946. The radio field was produced by a set operating on a wavelength of 3 metres at a height of 35 feet at Ashburton. Meteorological ascents made at different points in the path of transmission showed that the atmosphere was very nearly horizontally stratified, so that this assumption in the theory was justified.

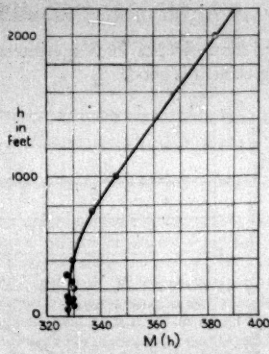


Figure 6. M-curve found 47 miles offshore from Ashburton, South Island, New Zealand, in late afternoon 4 November 1946.

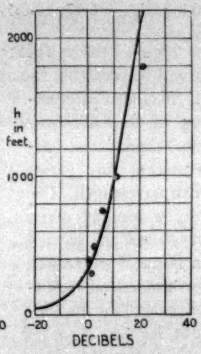


Figure 7. Height gain on $\lambda=3$ m. The dots represent measured values; the full line is obtained from single-mode wave theory.

The M-curve represents an interesting situation in which a duct just fails to occur; M is constant up to about 400 feet and then increases linearly. For this M-curve only the first mode should be important at a range of 47 miles below the horizon. The situation was therefore simple enough for single-mode wave theory to apply. The method of analysis follows the lines described above : the M-curve is represented by a linear term modified by two exponential terms : $M(h)=310+0{\cdot}0365h+46{\cdot}4e^{-0{\cdot}0046h}-27{\cdot}8e^{-0{\cdot}0069h}$. This function is plotted in figure 6 together with the experimental points, and it will be seen that it represents a satisfactory approximation. The eigenvalue D_1 is then calculated for a range of wavelengths (table 4).

Lastly, the height–gain function is calculated by numerical integration of the wave equation using the eigenvalue appropriate to $\lambda=3$ metres (table 5).

Table 4. The eigenvalues and the exponential attenuation rates as functions of the wavelength for the M-curve of figure 6

z		w	c		$310w+D_1=D/H$		γ	λ (cm.)
Re	Im		Re	Im	Re	Im		
0·4047	−0·2	0·5796	0·4134	0·2040	−10·933	0·152	0·130	16·1
0·5685	−0·4	0·2488	0·5710	0·4018	− 4·505	0·489	0·302	57·3
0·7106	−0·6	0·1551	0·6396	0·5401	− 2·666	0·817	0·398	116·3
0·8319	−0·8	0·1130	0·6674	0·6418	− 1·893	1·087	0·451	187·1
0·9360	−1·0	0·0895	0·6731	0·7191	− 1·507	1·309	0·484	265·4
1·0268	−1·2	0·0747	0·6654	0·7776	− 1·301	1·481	0·500	348·1
1·1077	−1·4	0·0645	0·6507	0·8226	− 1·190	1·615	0·507	433·8
					− 1·402	1·387	0·492	300

γ = exponential attenuation rate in db/mile.

Table 5. Theoretical height–gain function

h (feet)	159	318	445	636	764	954	1080	1272	1400	1590
20 log $\lvert U_1 \rvert$	−6·07	−0·17	2·72	5·87	7·57	9·77	11·09	12·92	14·03	15·62

Figure 7 shows a plot of the theoretical and experimental height–gain functions. It will be seen that the agreement is good.

It is also interesting to consider how the exponential attenuation rate varies with wavelength. The results already given in table 4 are plotted in figure 8 with those for a standard atmosphere and for the 350-foot surface duct of § 3(i). It will be seen that the attenuation rate is small, but not zero, even at very short wavelengths, which is accounted for by the fact that the M-curve fails to show any duct. This type of M-curve, with a vertical tangent at ground level and no duct, represents a limiting case which divides M-curves into two categories: those with ducts and those without ducts.

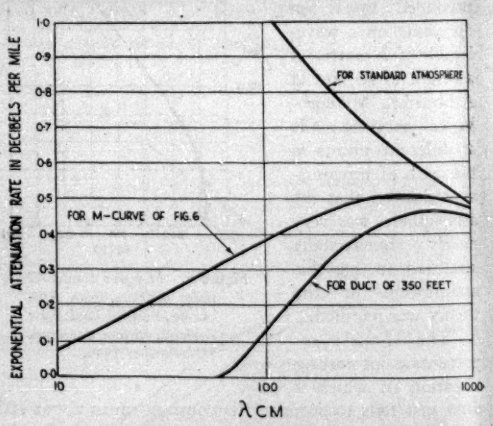


Figure 8. Dependence of exponential attenuation rate on wavelength for the M-curves of figures 1 and 2.

ACKNOWLEDGMENT

This work has been done at T.R.E., a Ministry of Supply Establishment, under contract to D.S.I.R. and as part of the Radio Research Board (D.S.I.R.) programme. It has been carried out in conjunction with D.S.I.R. New Zealand,

to whom thanks are due for permission to use the radio and meteorological data recorded at Ashburton, New Zealand, on 4 November 1946.

Crown copyright is reserved. The paper is reproduced with the permission of the Controller of H.M. Stationery Office.

REFERENCES

BOOKER, H. G., and WALKINSHAW, W., 1946, *Meteorological Factors in Radio-Wave Propagation* (London : Physical Society), p. 80.

DAVIS, H. T., 1935, *Tables of the Higher Mathematical Functions*, Vol. II (Indiana : Principia Press Inc.).

ECKERSLEY, T. L., and MILLINGTON, G., 1938, *Phil. Trans. Roy. Soc.* A, **237**, 273.

HARTREE, D. R., MICHEL, J. G. L., and NICOLSON, P., 1946, *Meteorological Factors in Radio-Wave Propagation* (London : Physical Society), p. 127.

HAY, H. G., 1948, *Phil. Mag.*, in press.

MACFARLANE, G. G., 1947, *Proc. Camb. Phil. Soc.*, **43**, 213.

PEKERIS, C. L., 1946, *J. Appl. Phys.*, **17**, 678.

Radio Shadow Effects Produced in the Atmosphere by Inversions

BY W. L. PRICE

Division of Radiophysics, Council for Scientific and Industrial Research, Commonwealth of Australia

MS. communicated by L. G. Dobbie ; received 21 *July* 1947

ABSTRACT. Reports from radar operators in Australia in 1940 revealed marked increase in range of surface objects on certain occasions apparently associated with distinctive meteorological conditions. It appeared likely that this phenomenon should be accompanied by the existence of " shadow " zones in the atmosphere, in which aircraft would not be detected. A theoretical investigation by ray-tracing methods shows that such shadow zones should exist above certain inversions. Controlled experiments, in which measurements of all the relevant physical conditions were made, show that these shadow zones do exist. The position and extent of the zones agrees well with the theory.

§ 1. INTRODUCTION

SHORTLY after the installation of coastal radar equipment on the east coast of New South Wales, in May 1940, operators began to report the occasional detection of surface objects at abnormally great ranges. The Commander Fixed Defences, Major-General J. S. Whitelaw, pointed out to Dr. D. F. Martyn that this effect coincided usually with the existence of an extensive anticyclone system over the eastern part of the Australian continent. Theoretical and experimental investigation was initiated, at first by the Radiophysics Laboratory of the Council for Scientific and Industrial Research, and later by the Australian Army Operational Research Group (O.R.G.).

It was early realized that the phenomenon was mainly due to the sharp temperature and humidity gradients associated with certain meteorological conditions, notably temperature inversions (Englund, Crawford and Mumford 1938), and attention was concentrated on this aspect Information soon reached

Australia that a similar investigation in England, initiated principally by Booker and Walkinshaw (1946), was directed mainly towards the determination of the increased fields close to the surface of the earth, in the "radio duct" formed within and below such gradients by the "trapping" of radiation. Now it seemed probable that under abnormal conditions of high surface fields there should occur a weakening of the field above the abnormal gradients. If sufficiently severe this might explain the occasional failure of aircraft warning radar sets to detect aircraft flying at certain heights when within the normal range of the set (Martyn 1943). Accordingly attention was devoted in Australia to this latter aspect, which was of considerable operational importance.

In the early stages of the investigation an attempt was made to derive information from the log books of radar operators and aircraft pilots. However, this method proved unreliable and a series of carefully controlled flights off the Sydney coast was made during the summer of 1942–3. Steep climbs were made by an aircraft at a range of about 24 miles from a 200 Mc/s. transmitter, 250 feet above mean sea level. Field strength and temperature were measured in the plane. Humidity was not measured, but use was made of hygrometer readings taken by the R.A.A.F. during their regular meteorological flights at Richmond, about 25 miles inland. At the same time propagation through a horizontally stratified medium was investigated theoretically by a geometrical optics treatment, this being considered best for the region near the level of inversion (Price 1943). The experimental results were published in a service report by D. R. Brown (1943). Examination of the results showed that on days of temperature inversion the three lowest lobes of radiation were depressed from their normal position, while higher lobes were not measurably affected. Analysis by ray methods showed that the extent of the lowering was of the right order (Price 1943). The results indicated that further experimental work was desirable, especially along the lines of combined air sounding and field strength measurements. Work was commenced on the development of accurate instruments for this purpose, and taken over by the Radiophysics Laboratory after the Operational Research Group was disbanded in 1944.

Upon the completion of sounding equipment designed for use in aircraft, a series of flights was carried out off the N.S.W. coast near Sydney in January 1946 by the Radiophysics Laboratory. The results proved of interest in connection with (*a*) the development of inclined stratified air masses produced by off-shore streaming of hot dry air, modified to some extent by a sea-breeze type of circulation, and (*b*) the sharp decrease of field strength with height experienced on passing through a layer with negative refractive index gradient.

This paper discusses the ray theory of propagation of ultra-high-frequency radio waves through such layers. The general theory for inclined layers developed in § 2 is applied in §§ 3 and 4 to an analysis of the field above the region of surface trapping caused by a negative refractive index gradient. Rays have been traced for one of the January flights and the results are compared in § 5 with radar observations of echo strength from the aircraft.

§ 2. RAY THEORY FOR A STRATIFIED ATMOSPHERE

2.1. *Scope of ray theory*

An essential feature of ray theory is that the radiation field can be described in terms of the propagation of energy along rays whose direction is given by

the laws of refraction, a given parcel of electromagnetic flux travelling within a tubular space bounded by the rays. Surfaces of equal phase, orthogonal to the rays, constitute the wave front. This description will be valid, provided that the laws of the electromagnetic field (i.e. Maxwell's equations) are satisfied. It is known that there are limitations to the validity of ray theory, the main ones being:—

(i) $|\text{grad}\,\mu| \ll 1/\lambda$, where λ is the wavelength. This means that the radius of curvature of the ray must be sufficiently greater than λ.

(ii) The rays must not come to a focus.

(iii) The region considered is not near the boundary rays produced by obstacles, nor within confined spaces whose dimensions are comparable with the wavelength.

To some extent ray theory can be extended by the use of Huyghens' Principle and the idea of interference. When this method becomes too complex other solutions of the wave equation are sometimes available which then enable the field to be described in simpler terms, such as, for example, by means of transverse electric (or magnetic) modes. Which is used depends on the nature of the problem and the completeness required in the solution. Treatment by ray methods, keeping their limitations in mind, sometimes provides a useful if approximate solution when a more complete quantitative answer is arithmetically, if not algebraically, impracticable. This appears to be the case when dealing with the propagation of U.H.F. radio waves over the earth, just above a refractive index inversion, in the region where the lobe pattern begins to be sensibly affected by super-refraction.

2.2. *Equation of the ray*

Figure 1 shows a ray TR from a transmitter T, at height h_T above the surface of the earth, which is assumed spherical with radius a, and centre O. The position of any point R is given by polar coordinates $(a+h,\ \theta)$ or by height h and reduced range d measured along the surface of the earth. The contours of equal refractive index, μ = constant, may be curved with respect to the earth and inclined at small angle to the horizontal. For simplicity, cylindrical symmetry about the vertical through the transmitter is assumed, so that the path of a ray lies in a vertical plane. In general, this would not be strictly accurate and lateral deviations (which could be important in bearing measurements) are possible, but since all deviations are small, it is considered that vertical and horizontal refraction can be treated independently.

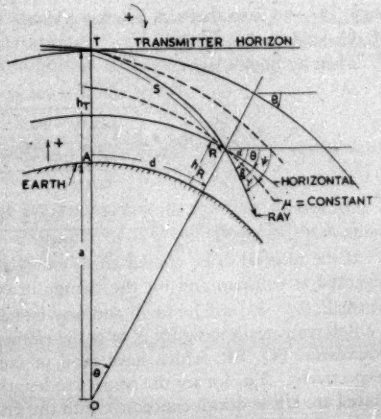


Figure 1. Coordinate system used in text for a ray passing through inclined layers.

We define the following symbols (figure 1):—

α = angle made by a ray to the horizontal at the transmitter $= \beta + \theta = \gamma + \psi$.

β = angle made by a ray to the horizontal.

γ = angle made by a ray to the stratification (angle of elevation).

ψ = angle made by the stratification to the horizontal at the transmitter.

θ = angle made by the horizontal to the horizontal at the transmitter.

r = radius of curvature of the stratification, μ = constant (figure 2).

q = radius of curvature of the normal to the stratification in the plane of the ray.

ρ = radius of curvature of the ray (figure 2).

s = distance measured along the ray.

Clockwise rotations are positive, so that downward curvatures are positive, in accordance with the usual convention for the earth's curvature.

Consider a ray passing from a point P, where the refractive index is μ, to R, where it is $\mu + \delta\mu$, while the complement of the angle of incidence changes from γ to $\gamma + \delta\gamma$ (figure 2). The deviation of the ray is $\delta\alpha = \delta\gamma + \delta\psi$.

As the ray passes from P to R the angle of incidence changes continuously, part of the change being due to the rotation of the contours. To the first order the effect of this part on the deviation of the ray can be separated from the effect of the alteration in γ due to refraction, which is taken care of in the formula by the changing value of μ, as in the case of parallel layers. Let the effective average change in the elevation angle be $-\overline{\delta\psi}/2$ so that the average value of this angle is $(\gamma - \overline{\delta\psi}/2)$.

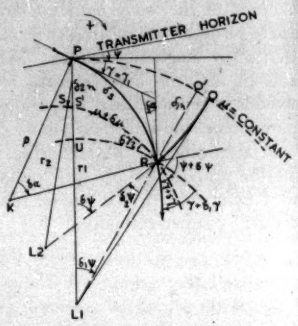


Figure 2. Deviation of an element of a ray produced by a thin stratum.

Then by Snell's law

$$\left.\begin{aligned}\frac{\delta\mu}{\mu} &= \frac{\cos(\gamma - \overline{\delta\psi}/2) - \cos(\gamma - \overline{\delta\psi}/2 + \delta\alpha)}{\cos(\gamma - \overline{\delta\psi}/2 + \delta\alpha)} \\ &\simeq \frac{\sin(\gamma - \overline{\delta\psi}.2 + \delta\alpha/2)\delta\alpha}{\cos\gamma} \\ &= \frac{\sin\{\gamma + \delta\gamma/2 - (\overline{\delta\psi}/2 - \delta\psi/2)\}\delta\alpha}{\cos\gamma}.\end{aligned}\right\} \quad \ldots\ldots(1)$$

If the interval δs be chosen small enough for each contour curvature to be regarded as constant and for the change in curvature between the contours to be small, $(\overline{\delta\psi} - \delta\psi)$ will be small and may be neglected.

Referring again to figure 2, δs is the element of the ray between the stratum boundaries PQ, SR, which have radii of curvature r_1, r_2 and centres L_1, L_2 respectively. $\delta_1 n$, $\delta_2 n$ are the respective lengths of the normals Q'R, PS'. The dotted arc UR is drawn concentric with the first surface PQ.

$\delta\gamma$, the deviation of the ray relative to the strata can be considered to consist of two parts: $\delta_1\gamma$, the rotation of the ray relative to the first surface, in going

from P to R; $\delta_2\gamma$, equal to the angular displacement at R of the first surface relative to the second. $\delta\psi$ may be similarly subdivided. Then

$$\delta\gamma=\delta_1\gamma+\delta_2\gamma,\ \delta\psi=\delta_1\psi+\delta_2\psi,\ \delta_2\gamma=-\delta_2\psi,\ \delta\alpha=\delta\gamma+\delta\psi=\delta_1\gamma+\delta_1\psi. \quad \ldots\ldots(2)$$

$\delta_2\psi$ is due to the displacement of the centre of curvature.

Equation (1) may be written

$$\cos\gamma\frac{\delta\mu}{\mu\delta_1 n}=\sin(\gamma+\delta_1\gamma/2)\frac{\delta\alpha}{\delta_1 n}+\{\sin(\gamma+\delta\gamma/2)-\sin(\gamma+\delta_1\gamma/2)\}\frac{\delta\alpha}{\delta_1 n},$$

i.e.
$$\frac{\cos\gamma(\delta\mu/\mu+\delta_2\psi\,.\,\delta\alpha/2)}{\delta_1 n}\simeq\sin(\gamma+\delta_1\gamma/2)\frac{\delta\alpha}{\delta_1 n}=-1/\rho. \quad \ldots\ldots(3)$$

since
$$\cos(\gamma+\delta\gamma/4+\delta_1\gamma/4)\simeq\cos\gamma,$$

$1/\rho$ is here the average curvature of the ray PR.

Equation (3) may be rearranged in the following form, suitable for computation:

$$\cos\gamma\frac{\delta\mu}{\mu\delta_1 n}\left(\frac{1}{1-\cos\gamma(\delta_2\psi/2)/\sin(\gamma+\delta_1\gamma/2)}\right)=\sin(\gamma+\delta_1\gamma/2)\frac{\delta\alpha}{\delta_1 n}. \quad \ldots\ldots(4)$$

From the geometry of PQ'R, to the first order of small quantities,

$$\cos\gamma/r_1=-\sin(\gamma+\delta_1\gamma/2)\delta_1\psi/\delta_1 n, \quad \ldots\ldots(5)$$

remembering that δn is positive when directed upwards and r_1 is positive when the surface is concave downward. From equations (4) and (5), noting that γ_1 can be put equal to γ, we get the following relations:

$$\cos\gamma_1\left\{\frac{\delta\mu}{\mu\delta_1 n}\left(\frac{1}{1-\cos\gamma_1(\delta_2\psi/2)/\sin(\gamma_1+\delta_1\gamma/2)}\right)+\frac{1}{r_1}\right\}=\sin(\gamma_1+\delta_1\gamma/2)\frac{\delta_1\gamma}{\delta_1 n} \quad \ldots\ldots(6)$$

$$=\cos\gamma_1(1/r_1-1/\cos\gamma_1\rho)=\cos\gamma_1/k_1, \quad \ldots\ldots(7)$$

which is the average curvature of stratification at P relative to a ray inclined to it at γ. $1/k_1$* can be called the average "effective curvature" (Schelleng, Burrows and Ferrell 1933) of the stratification at P.

When $\gamma\ll 1$ and $\mu\simeq 1$ we get

$$\frac{\delta_1 n}{k_1}=\delta\mu\left(\frac{1}{1-(\delta_2\psi/2)/(\gamma_1+\delta_1\gamma/2)}\right)+\frac{\delta_1 n}{n_1}=(\gamma_1+\delta_1\gamma/2)\delta_1\gamma=\delta_1(\gamma^2/2), \quad \ldots\ldots(8)$$

$(\gamma_1+\delta_2\gamma/2)$ is the average slope of the ray relative to the first surface and is approximately equal to $-\delta_1 n/\delta_1 d$. $(\delta_2\psi/2)(\gamma_1+\delta_1\gamma/2)$ can always be made small compared with unity by decreasing $\delta\mu$ sufficiently.

Since it can be shown that

$$k_1/r_1=-\delta_1\psi/\delta_1\gamma \quad \ldots\ldots(9)$$

from (5), (6) and (7), it follows that the range measured along the stratification at P will be given by

$$\delta_1 d=r_1\delta_1\psi=-k_1\delta_1\gamma. \quad \ldots\ldots(10)$$

* Sometimes the symbol k is used for the ratio ("effective radius"/actual earth's radius), instead of for effective radius, as here.

The optical distance is

$$(\mu+\delta\mu/2)\cos(\gamma_1+\delta_1\gamma/2)\delta_1 d-(\mu+\delta\mu/2)\sin(\gamma_1+\delta_1\gamma/2)\delta_1 n \quad \ldots\ldots(11)$$

and so can be separated into two components, one depending upon the distance along the stratum and the other along the normal. $-\delta_2\psi/\delta_1 n$, which is a measure of the curvature $1/q$ of RQ, the line orthogonal to the strata, and the curvatures $1/r_1$, $1/\rho$, must be finite or zero for all values of γ as $\delta\mu$, $\delta\alpha$ etc. approach zero, unless there is a discontinuity in the slope of the contours. Thus, in the limit, equations (6) and (7) reduce to

$$\frac{\cos\gamma}{k}=\cos\gamma\left(\frac{d\mu}{\mu dn}+\frac{1}{r}\right)=\sin\gamma\frac{d_1\gamma}{dn}=\sin\gamma\left(\frac{d\gamma}{dn}-\frac{1}{q}\right) \quad \ldots\ldots(12)$$

when $\delta\mu\to 0$. Also

$$1/k=(1/r+d\mu/\mu\, dn)=(1/r-1/\rho_T), \quad \ldots\ldots(13)$$

$$d_1 d=r d_1\psi=-k d_1\gamma. \quad \ldots\ldots(14)$$

In the limit the effective curvature of the stratification, $1/k$, becomes equal to the curvature of this surface, relative to the tangent ray at P, whose curvature is

$$1/\rho_T=-(1/\mu)(d\mu/dn).$$

Numerical computation of the ray path. When γ is small equations (8) and (10), giving finite differences, can be used for tracing the path of the ray. If a numerical procedure is adopted the path of the ray is determined in steps between consecutive surfaces $\mu=$ constant, spaced so that the curvatures of each surface and of the ray are approximately constant, and the changes in their slopes ($\delta\psi$, $\delta\alpha$) are small. $\delta_2 n$, $1/r_1$, $1/r_2$, γ_1 and $\{\delta_2\psi-(\delta d/r_2-\delta d/r_1)\}$, quantities which occur at the beginning of a step, can be read at each stage from a large scale plot of the contours, similar to figure 8 (*a*). The last quantity is equal to the initial difference in ψ between the two surfaces. An approximate value of δd can be obtained by inspection of the previous path of the ray.

Values of $\delta_1 n$, $\delta_2\psi$ and $(\gamma+\delta_1\gamma/2)$ are then calculated from the relation

$$(\gamma+\delta_1\gamma/2)=-\delta_2 n/\delta d+\{\delta_2\psi-(\delta d/r_2-\delta d/r_1)+\tfrac{1}{2}(\delta d/r_2-\delta d/r_1)\}=-\delta_1 n/\delta d. \quad \ldots\ldots(15)$$

Using these values, (8) is solved for $\delta_1\gamma$ by successive approximation. With the new value of $(\gamma+\delta_1\gamma/2)$, (15) is solved for δd leading to new values of $\delta_2\psi$ and $\delta_1 n$. The process is repeated in (8) and (15) until consistent values of $\delta_1\gamma$ and $\delta_1 d$ are obtained. Equation (10) should then check automatically.

It is convenient to use the modified refractive index described in § 2.3.2.

2.3. *Transformation of coordinates*

The form of expressions (12), (13) and (14) shows that the relative curvature of the ray, at any point, viz. $-(\cos\gamma)/k$, is derivable from a "modified" refractive index N whose logarithmic gradient dN/Ndn is equal to $1/k$, the stratum at the point being regarded as flat. We may also regard the process as one of subtracting a curvature $(\cos\gamma)/r$ from any ray making an angle γ to the stratification. Inclination of the contours due to change in the position of the centre of curvature is not thereby eliminated.

2.3.1. *Concentric stratification.* When the strata are concentric (13) becomes

$$1/k=d\mu/\mu dr+1/r=\tan\gamma d\gamma/dr, \quad \ldots\ldots(16)$$

since now $d\gamma = d_1\gamma$; $d_2\gamma = 0$; $d_1 n = dr$. Integrating, we get the well known equation of the ray

$$\mu r \cos\gamma = \text{constant}. \qquad \ldots\ldots(17)$$

2.3.2. *Horizontal stratification.* Further, if the stratification is horizontal

$$1/k = d\mu/(\mu dh) + 1/(a+h) = dN/(Ndh) = \tan\gamma d\gamma/dh \qquad \ldots\ldots(18)$$
$$\simeq d(\gamma^2/2)/dh, \quad \gamma \ll 1,$$

and
$$dd = ad\psi = -akd\gamma/(a+h),$$

where d is here the range, reduced to the earth's surface.

Integrating (18) with $N = \mu^*$ when $h = 0$ we get

$$N - 1 = \mu(a+h)/a - 1 = (\mu - 1) + h/a + (\mu - 1)h/a \qquad \ldots\ldots(19)$$

and
$$N\cos\gamma = \text{constant}. \qquad \ldots\ldots(20)$$

The following approximation to the "modified" index N is generally used (Freehafer 1943):

$$N - 1 = (\mu - 1) + h/a, \ \mu \simeq 1, \ h \ll a. \qquad \ldots\ldots(21)$$

Also it is customary to define a quantity M, of convenient size, by

$$10^6 M = (\mu - 1) + h/a. \qquad \ldots\ldots(22)$$

2.3.3. *Plot in rectangular coordinates* (Booker and Walkinshaw 1946). It follows from the last section that, in a horizontally stratified medium situated over a flat earth and having a refractive index N given by (18) or (19), a ray will have the same relation between γ, h and reduced range d, as in a horizontal curved medium, refractive index μ, over a curved earth of radius a. Thus a plot in rectangular coordinates of reduced range against height will give the same ray paths as would be obtained by straightening the strata and making their length equal to their projection to ground level. These will not be quite the same curves as obtained using N over a plane earth, unless the ranges in the latter case are also reduced by the factor $a/(a+h)$.

It is convenient to eliminate the curvature of the earth and use a modified index of refraction even in the case of non-horizontal stratification. When h/a is now added to μ to obtain N, the lines of constant N will not remain parallel to those of constant μ, but become inclined at an angle ϵ, given by

$$(-d\mu/d_1 n)\sin\epsilon = (1/a)\sin(\psi - \theta + \epsilon), \qquad \ldots\ldots(23)$$

this being the condition for zero gradient of N in the direction inclined at ϵ to the lines $\mu = \text{constant}$ (figure 3).

It will be shown that the path of the ray will be approximately the same when N is taken as the refractive index and the earth's curvature disregarded as when the actual refractive index and curvature are used, provided the angles of inclination are small.

In figure 3 the angles and radii are defined as before, dashes distinguishing those measured to

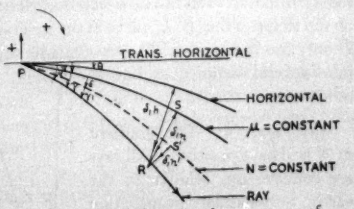


Figure 3. Transformation from contours of μ = constant to those of N = constant.

* The unnormalized value of N is here defined.

N = constant. Cosines of small angles can be put equal to 1. We have from equation (1)

$$\frac{d\mu}{d_1 n} = \tan\gamma \frac{d\alpha}{d_1 n} \quad (\mu \simeq 1) \quad = \frac{\cos\gamma'}{\cos\gamma} \tan\gamma' \frac{d\alpha}{d_1 n'}$$

since angles RSP, RS'P each equals 90°.

Therefore $$\left(\frac{d\mu}{d_1 n}\cos\epsilon + \frac{\cos(\psi - \theta + \epsilon)}{a}\right) - \frac{\cos(\psi - \theta + \epsilon)}{a} = \cos\epsilon \frac{\cos\gamma'}{\cos\gamma} \tan\gamma' \frac{d\alpha}{d_1 n'}.$$

But $$1/r_N = -\tan\gamma' \frac{d_1(\psi + \epsilon)}{d_1 n'} = \text{curvature of PS'};$$

on putting $d\alpha = d(\psi + \epsilon + \gamma')$ and adding the last two equations we get

$$\left(\frac{d\mu}{d_1 n}\cos\epsilon + \frac{\cos(\psi - \theta + \epsilon)}{a}\right) + \left(\frac{1}{r_N} - \frac{\cos(\psi - \theta + \epsilon)}{a}\right) \simeq \tan\gamma' \frac{d_1\gamma'}{d_1 n'}$$

or $$dN/d_1 n' + 1/r' = \tan\gamma' d_1\gamma'/d_1 n' \qquad \ldots\ldots (24)$$

since $1/r' = 1/r_N - \cos(\psi - \theta + \epsilon)/a$ = curvature of the N stratum relative to the earth and $((d\mu/d_1 n)\cos\epsilon + \cos(\psi - \theta + \epsilon)/a)$ is the resultant gradient of N. Equation (24) gives the same form for the variation of γ with height as when N is the refractive index and the earth is flat.

§ 3. CHARACTERISTIC OF N AND M CURVES FOR HORIZONTAL STRATIFICATIONS

An advantage of the use of a modified index lies in showing quickly how the inclination of a ray to the horizontal changes with height. When the gradient of N (or M) is positive, γ continually decreases with height and the ray is convex upwards relative to the earth; conversely for a negative gradient. The usual condition of the atmosphere is for grad N to have a positive value lying between 30×10^{-9}/ft. and 45×10^{-9}/ft. and increasing with height. The standard value of $1/k$ generally used at low levels is 36×10^{-9}/ft. Temperature inversions and, more especially, abnormal hydrolapses producing negative gradients of μ do occur quite frequently, but they can only persist over limited intervals of height. The resulting structure of the atmosphere shows a series of maxima and minima of N with height, the greatest and most important N deficits occurring below the first one or two thousand feet. Higher level N inversions which occur can reflect radiation to an extent which is sometimes significant, the amount depending on the shape of the (N, h) curve at the inversion, and on the propagation frequency. If only the lower inversions are considered, it is found that the (N, h) curves fall into several definite classes as follows:—

(a) *Figure* 4. The gradient of N is always positive. All rays are concave up, relative to the earth. Sub-standard propagation occurs when the value of the gradient of N is greater than normal (36×10^{-9}/ft.).

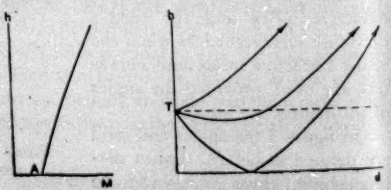


Figure 4. Paths of a ray when gradient M is positive.

(b) *Figure* 5. The initial gradient is negative and a minimum occurs at B. Rays are concave down or up, according as they are below or above B.

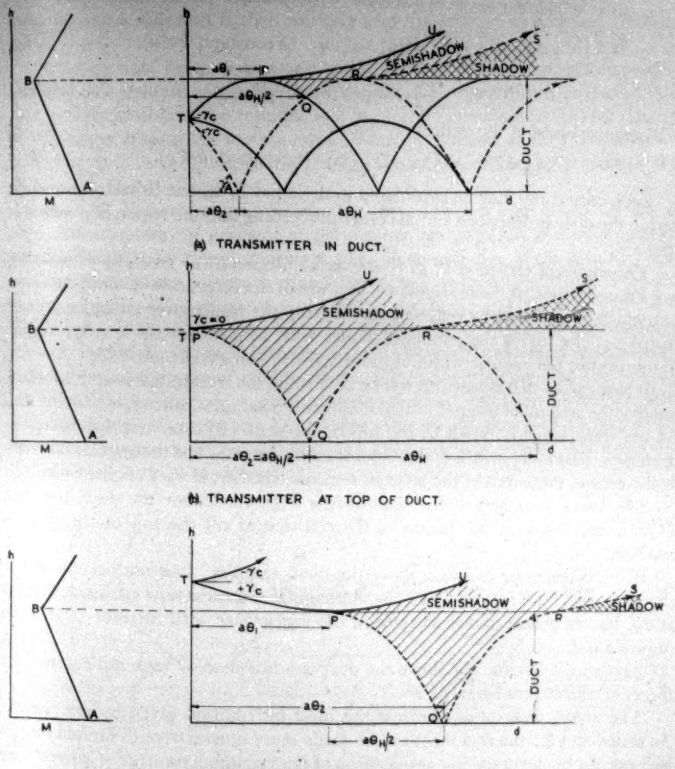


Figure 5. Shadow regions caused by a surface duct.

From a transmitter below B (figure 5 (*a*)) there will be a pair of rays leaving at angles $\pm\gamma_c$, which just reach the level of the minimum of N, that at $+\gamma_c$ doing so after reflection from the earth. The flux emitted between these rays is trapped between the earth and the horizontal through B, this region being called a surface duct (Booker and Walkinshaw 1946).

The upper critical ray, initial angle $-\gamma_c$, reaches the level of B at a finite distance, whose angular range θ_1 is given approximately by

$$\theta_1 = \int_{-\gamma_c}^{0} -kd\gamma/(a+h) \simeq -k\gamma_c/a; \quad (k \text{ constant}, h \ll a), \quad \ldots\ldots(25)$$

$$\gamma_c = +\sqrt{2(N_T - N_B)}. \quad \ldots\ldots(26)$$

Then it returns to earth in a half-hop, is reflected, and continues between earth and level B in full hops around the earth. The lower critical ray, $+\gamma_c$, reaches the earth at an angular range θ_2 given by

$$\theta_2 = \int_{\gamma_c}^{\gamma_A} -k d\gamma/(a+h) \simeq -k(\gamma_A - \gamma_c)/a; \quad (k \text{ constant}, h \ll a). \qquad \ldots\ldots(27)$$

$$\gamma_A = +\sqrt{2}(N_A - N_B)^{\frac{1}{2}}. \qquad \ldots\ldots(28)$$

The angular range of the critical ray hop is

$$\theta_H = 2(\theta_1 + \theta_2) \simeq -2k\gamma_A/a; \quad (k \text{ constant}, h \ll a). \qquad \ldots\ldots(29)$$

Each critical ray may be considered to divide at P, R, one branch continuing within the duct as described and the other proceeding into the region of positive k, above B.

The angles of elevation at P, R are zero. Between the two upper branches is a laminar region of "semi-shadow" into which the direct rays do not penetrate, but which is traversed by reflected rays leaving the transmitter at angles greater than γ_c. The tilt of the lamina increases with height, but for a given height of transmitter the angular width of the shadow remains constant at $2\theta_2$.

It is apparent that this width will be zero when the transmitter is at the surface but that the amount of trapped radiation is then a maximum, since γ_c is a maximum. As the transmitter is raised to the top of the duct (B), the trapped radiation decreases to zero, the width of the shadow increases to θ_H but the angular distance to the centre, measured at the level B, remains constant at $\theta_H/2$ (figure 5 (*b*)).

The lower boundary of the semishadow PQR lies below the duct, the area PQR increasing with the height of the transmitter till the top of the duct is reached.

If the transmitter is raised above the duct, trapping does not occur but the shadow zone is still formed. The lamina angular thickness now remains constant at θ_H, but its centre moves away from the transmitter with increase of altitude (figure 5 (*c*)).

Beyond the lamina and above the duct is a full shadow zone not reached by direct or reflected radiation.

The above description presents the ideal distribution given by ray theory. As stated in §2.1 the treatment can be made more quantitative to limited ranges determined by the practical consideration of the maximum number of intersecting rays which can be handled by interference methods. At long ranges, within the duct, use has been made of the theory of guided waves, the minimum range to which the field may be calculated being determined in this case by the maximum number of modes which can conveniently be dealt with. Considerable work has been done in this way on surface trapping. Intermediate ranges are not amenable to either method and little work by way of calculation has been done for the semishadow region.

It is clear that some radiation will be diffracted into the "semishadow" and "shadow" across the curved surfaces PU and RS, and that some will penetrate the boundary PQR, in accordance with wave theory, but the diffracted field will generally be weak compared with the radiation which would reach the space in the absence of trapping. Neglecting then the diffracted field in order to obtain an upper limit to the decrease in intensity, it is seen that the normal interference field at the maxima of the lobes will be decreased about 6 db. in the semishadow

region, due to the absence of the direct ray. (This does not take into account increased divergence of the reflected ray due to the negative gradient.) Such a decrease would be important when searching for aircraft at long and medium ranges, since a radar set is then working near the limit of its ordinary detection range.

Ray treatment predicts a second major decrease in field strength on passing into the full shadow across the false horizon RS—theoretically to zero. The problem of calculating the residual field beyond RS is similar to that for the case of diffraction around the earth's surface near the horizon when superrefraction is absent, but is complicated by leakage across the diffracting boundary RV. A full wave theory is required.

The (M, h) curves upon which the discussion has been based are drawn with discontinuities of gradient at the maxima and minima, and with constant gradient between. Actual soundings show some rounding of the turning points and curvature of the joining sections, but as will be seen from figure 8 (b), this is often small. The effect of rounding of the minima is to make grad $N \to 0$, and $d \to \infty$. The critical rays tend asymptotically to approach the horizontal through B, but the ray for which the value of γ at level T is slightly greater than γ_c cuts the top of the duct at a finite range. The region bounded by the two rays at $\pm\gamma$ will only contain a small amount of direct radiation above the duct, proportional to $\gamma - \gamma_c$, and may be regarded as the semishadow zone (cf. figure 8 (c)). One result is that the shadow boundaries will be diffuse instead of sharp, a condition which is aided by diffraction.

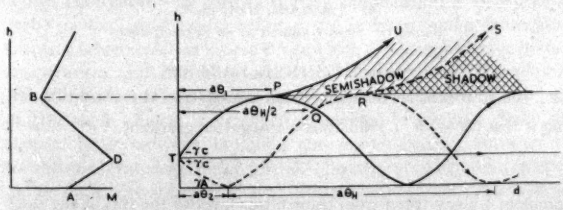


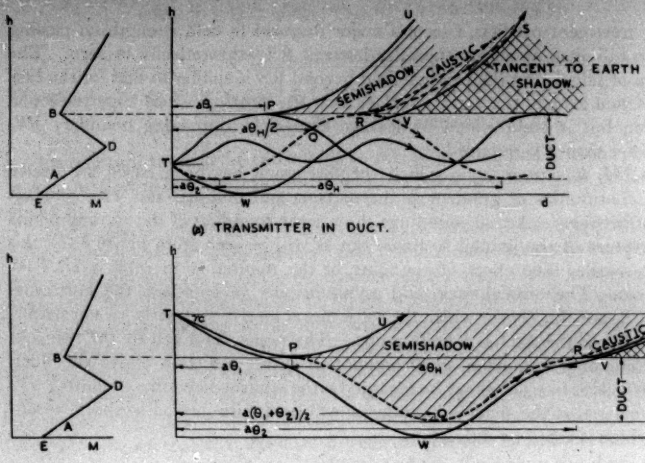
Figure 6. Surface duct. Elevated layer. Transmitter at max. M.

(c) *Figure* 6. The initial gradient is positive and is followed by the duct-forming negative gradient with a minimum N less than the surface value. The rays below D will be concave up, and maximum trapping occurs with the transmitter at D. Conditions otherwise vary much as in (b).

(d) *Figure* 7. As in (c), a positive gradient precedes a negative gradient, but the minimum N now is greater than the surface value. This is an elevated duct.

Here the lower critical ray descends only to the level A, where N has the same value as at B, so that the duct lies between the horizontals A and B. Interesting features are (i) the caustic passing through R, extending down to the level of the transmitter and up into the shadow; and (ii) the extension of semishadow beyond R, the full shadow boundary now being the continuation of the ray tangent to the earth beyond V. The rays whose intersections form the caustic lie between this tangent ray and the ray which leaves the transmitter horizontally.

When the transmitter is below the duct all radiation passes through it, so that there is no trapping or shadow. As the transmitter moves up through the duct, zero trapping occurs at A as well as at B, maximum trapping at D. The sequence of variation of shadow width and centre are similar to (b) and (c).



(*b*) Transmitter above duct.

Figure 7. Shadow caused by an elevated duct.

§4. INCLINED LAYERS

When the stratification is inclined, but concentric, the condition for possible trapping is that $(N+\int dr'/r')$ must have a negative gradient. We now have

$$\gamma_c' = \sqrt{\{-2\delta(N+\int dr'/r')\}},$$

the increment δ being taken from transmitter level to the top of the duct.

If the position of the centre of curvature is changing, the negative gradient of $(N+\int d_1 n'/r')$ must produce a change in elevation angle $\delta_1\gamma'$ greater than the rotation $\delta_2\gamma'$ due to the shift in curvature, in order that trapping may occur. In this case there is the possibility that rays trapped at one section may later penetrate the trapping layer owing to a change in its tilt. Again, rays initially above the minimum of $(N+\int d_1 n'_1/r')$ may be trapped on account of a favourable change in the tilt.

§5. EXPERIMENTAL FLIGHTS IN THE SUMMER OF 1946

5.1. *Locality and equipment*

In January 1946, a series of meteorological soundings were made in combination with radar measurements on a 200-Mc/s. C.O.L. Mark V air warning set at the radar station at Collaroy, 400 ft. above sea level and 15 miles north of the Sydney Heads. The aircraft flew about 100 miles out to sea at right angles to the coast and back, in a sequence of ascents and descents, to heights of 2000 to

3000 feet. On three of these flights observations were made on the radar set of the time, and of the range, bearing and signal-to-noise ratio of the echo from the plane. Observations of range were made every 10 seconds and of bearing every 2–3 minutes. The power output was monitored but an absolute calibration was not attempted.

In the aircraft a simultaneous record was made of time, number of the reading, air speed by air speed indicator, height, and hence pressure from an altimeter, and wet and dry wick temperatures, using thermocouples. The indicating meters were mounted on one panel and photographed at controllable intervals (3 to 6 seconds) by a camera with motor-driven shutter. The thermocouple and camera were designed and built by M. Iliffe and will be described elsewhere (Iliffe 1948). The hygrometer was mounted forward of the cabin and above the fuselage clear of slipstream disturbance and engine contamination. Times were synchronized with the radar station. Air speed and rates of climb were maintained as constant as possible by the pilot.

The flights were made on days forecast as favourable to strong M-inversions. At this time of the year the system of eastward moving anticyclones, which form the central motif of the Australian weather pattern, has a general track which crosses the eastern coast about latitude 40° S. and reaches the latitude of Sydney in the central Tasman Sea. The circulation associated with the rear portion of the anticyclone after it crosses the coast, creates strong off-shore streaming of very hot dry air. Modification by the cooler sea from below tends to produce the familiar surface advection duct with strong and deep temperature and humidity inversions. A further diurnal modification is produced on the circulation by the hot land–cool sea influence on the density and pressure gradients. This results in a large-scale sea breeze effect which produces a change in the shape of the temperature and humidity contours. The reversal of wind direction near the sea allows the surface layer of modified air to become cooler and moister, and more thorough mixing is achieved to greater height and closer inshore. The transition layer between modified and unmodified air becomes thinner and higher and an elevated duct is formed. The contours of constant temperature and humidity slope upwards out to sea, with a gradient which increases slightly from the land, producing an upward curvature. A further effect of the reversal of low level circulation is to slow down the eastward movement of the anticyclone, causing the inversion conditions to persist for some days.

The detailed history of the development of these "coastal fronts" (Martyn and Squires 1944) is of considerable interest, and this aspect is to be investigated further by the Radiophysics Laboratory. In this paper, however, the interest is in the illustration afforded of the decrease in field strength across the transition layer between the lower and upper air masses. The elevated radio ducts so formed have M-inversion layers of exceptional magnitude and steepness, changes in M of 30 units sometimes occurring in 150 feet.

5.2. *Experimental results of the soundings*

5.2.1. *Experimental results* for two flights made on the 10th and 11th January, are presented in figures 8 (*a*), 8 (*b*), and figures 9 (*a*), 9 (*b*) respectively and the ray analysis for the first flight in figures 8 (*c*), 8 (*d*). Figures 8 (*a*) and 9 (*a*) give the contours of constant M for the outward run, and figures 8 (*b*) and 9 (*b*) give (M, h) curves at each 10 miles of this path. Thus (*a*) and (*b*) present the vertical

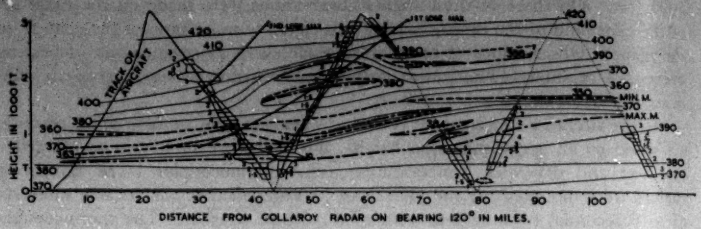


Figure 8 (*a*). Contours of M=constant for aircraft sounding, outward flight 10 Jan. 1946, from Collaroy, Sidney, N.S.W.

Lines of constant M ——— 360. Lines of maximum M — - — - Lines of minimum M - - - - -

The values of S/N are written alongside the track of the aircraft and also indicated by the distance apart of the lines drawn on each side of it. Where the aircraft is not seen the track is shown dotted thus .....

T=position of the radar set at Collaroy.

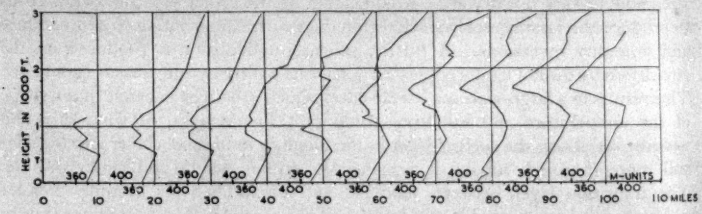


Figure 8 (*b*). M, h curves at a number of ranges for the stratification of figure 8 (*a*).

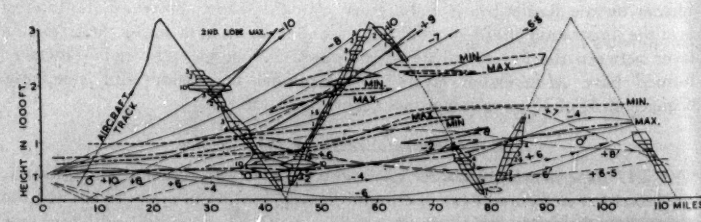


Figure 8 (*c*). Calculated ray paths from Collaroy radar for the stratification of figure 8 (*a*). Direct ray leaving transmitter at given angle of elevation ($\times 10^{-3}$ radians) -2 ; reflected ray $+2$.

Lines of maximum M — - — - Lines of minimum M - - - - - - S/N shown as in figure 8 (*a*).

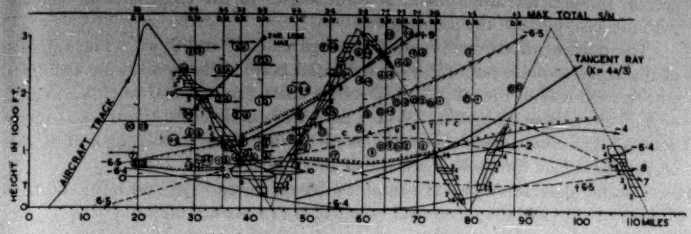


Figure 8 (*d*). Values of S/N calculated from divergence of rays in figure 8 (*c*).

Figures in circles are calculated values of average S/N at range of height lines beside which they are written ; average calculated for the height between the short horizontal lines. These values are represented by the horizontal lines terminated by arrow heads to the same scale as the measured S/N ; height lines are dotted when total $S/N < 1$. Figures at top of height lines are twice the calculated free space S/N for the given range.

Top of semishadow Bottom of semishadow

Top of shadow Bottom of shadow

D = direct, R = reflected.

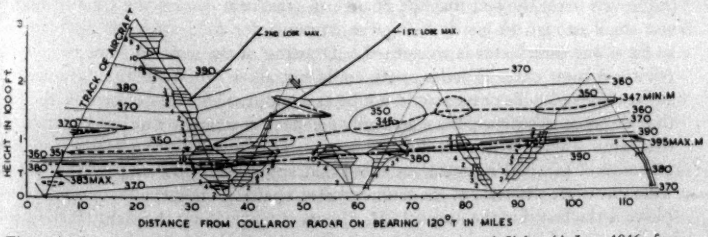


Figure 9 (*a*). Contours of M = constant for aircraft sounding, outward flight, 11 Jan. 1946, from Collaroy, Sydney, N.S.W. For details see figure 8 (*a*).

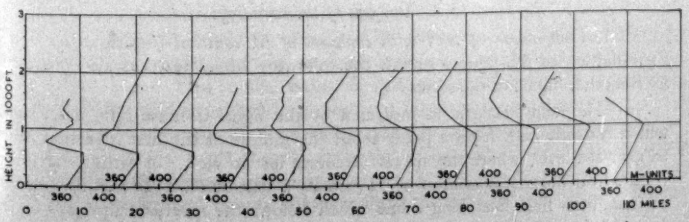


Figure 9 (*b*). M, h curves at a number of ranges from the stratification of figure 9 (*a*).

section through the atmosphere, at right angles to the coast, which coincides with the axial plane of the transmitter.

The track of the aircraft is shown in figures 8 (*a*) and 9 (*a*). The variation of the signal-to-noise ratio (S/N) is indicated graphically by drawing two lines, one on each side of the track, at a distance apart proportional to S/N. Values of S/N are written alongside. Where no echo was detected the track is shown dotted. For comparison, the positions of the first and second theoretical lines of maximum field for $k=4a/3$ are also drawn.

The ray diagram 8 (*c*) was computed from figures 8 (*a*) and 8 (*b*) and from this were calculated the theoretical values of S/N shown in figure 8 (*d*).

5.2.2. *Accuracy of the data.* The values of $(\mu-1)$ were calculated from Stickland's (1942) formula:—

$$\mu-1=(79P/T-11e/T+3{\cdot}8\times10^5e/T^2)10^{-6} \qquad \ldots\ldots(30)$$

the constants being rounded off to these values. P is the total pressure of the air, e the partial pressure of the water vapour (pressures in millibars) and $T°$ K. the absolute temperature of the air.

Temperatures were read to $\pm0{\cdot}1°$ C. and corrected for plane speed. Apart from the question of the accuracy of the formula constants, it is considered that the overall accuracy in M is ±1 M-unit, so far as individual readings are concerned. Accuracy of the contours is affected by the time factor involved in the duration of a flight and by the accuracy of the determination of position. The outward journey took about 50 minutes. A comparison made with the return flight (not given here) showed that the change in structure during the time of flights was small enough to justify using the contours for field strength calculations, so far as the time factor is concerned. Drawing of the contour lines by joining points, in some cases 20 miles apart, could not allow for small scale undulations and undoubtedly the curves could not be as smooth as shown. However, the fact that three consecutive points generally lie near a straight line would indicate that large scale undulations did not occur. Curves (*b*) were plotted from the contours of (*a*), and so again are somewhat smoothed. Bearing on this point, which involves the scale of the temporal and spatial homogeneity of the atmosphere is the fact that the original (M, h) curves taken along the slant track of the plane were, in fact, comparatively smooth. At the lower levels readings were usually made at height intervals of 15–20 feet and departure from the degree of smoothing represented by (*b*) seldom exceeded the accuracy of the M values, ±1 M-units. A point of interest in the curves (*b*) is the sharpness of the turning points where the gradients of M change sign.

5.3. *Analyses of the soundings*

5.3.1. *Comparison of S/N with contours of M constant* (figures 8 (*a*), 9 (*a*)). Examination of the figures reveals the following broad features also possessed by the other flights of the series:—

(*a*) The field strength, as indicated by the signal-to-noise ratio has a value which remains high from a point about the middle of the first inversion down to near sea-level, where the aircraft becomes lost to view. A strong maximum occurs near the first maximum of M (usually a little below it) so that the position of maximum field lies along a line which follows the inversion upwards as the latter rises out to sea.

(*b*) The field falls off rapidly shortly after the inversion is entered from below, sometimes to a value below receiver threshold, through a height interval which increases with range; above this S/N increases with height. This is shown by the values of S/N obtained during the second ascent and the first and second descents of the plane. The later climbs were not high enough to show the final increase of field. The upper boundary of the region of diminished field is thus an upward sloping line whose location is shown in §5.3.2 to be near the upper critical rays. The lower boundary is within the M inversion.

(*c*) The first lobe has disappeared in figure 9 (*a*). In figure 8 (*a*) there appears what could be an interference maximum representing the first lobe displaced upwards about 150 feet from its position with $k=4a/3$, and another at 57 miles 2650 feet, this time displaced upwards 600 feet. A neighbouring maximum at 55 miles would appear to be due to passage across the shadow boundary at this point.

(*d*) Above the upper boundary the field tends to resume its normal lobe structure as indicated by maxima at 30 miles 2050 feet in figure 8 (*a*) and 29 miles 1650 feet in figure 9 (*a*), representing the second lobe. In figure 8 (*a*) the position of the second lobe, like that of the first lobe, is high, while in figure 9 (*a*) the second lobe appears in its normal position.

5.3.2. *Comparison of S/N with ray tracing.* Rays were traced for the first sounding, covering the range of angles of elevation measured at the transmitter from 10×10^{-3} radians to -10×10^{-3} radians. These are plotted in figure 8 (*c*). The direct rays (negative α) are shown as full lines and the reflected rays (positive α) as dotted lines. The rays will be referred to by their initial angles of elevation in units of 10^{-3} radians. The lines of maximum and minimum M and the variation of S/N along the aircraft track are repeated from figure 8 (*a*).

It will be noticed that the elevated duct of figures 8 (*a*), 8 (*b*), is double, to a range of about 75 miles. The first M inversion, which falls to a deep minimum, is the dominant factor affecting the ray paths. This inversion is followed by a smaller duct whose influence is mainly noticed on the reflected rays. There is a further complication in the M-structure between 40 and 90 miles commencing at 2000 feet where another M-inversion occurs. On account of its height this has only a minor effect.

Both the direct and the reflected rays are strongly trapped by the first inversion. Splitting of the direct radiation occurs at about 11 miles 700 feet, the $-6{\cdot}5$ ray penetrating all the inversions while the $-6{\cdot}4$ ray is trapped by the first inversion (see figure 8 (*d*)).

The critical rays are not so clearly defined for the reflected radiation. The first inversion separates the $+6{\cdot}5$ ray from the $+7$ ray at about 40 miles 850 feet, and the critical ray would be about $+6{\cdot}8$. However, rays up to $+8$ become trapped by the second inversion, trapping being aided by the rise in the position of the inversion. This results in the full shadow being divided into two regions of diminished field, separated by a caustic. The upper region extends to about the $+9$ ray, but overlaps the semishadow, the two top edges crossing at 36 miles.

An estimate was made of the limits of the field strength values above the duct in terms of S/N, from a consideration of the divergence of the rays. It is felt that the experimental figures obtained from the radar measurements did not warrant closer calculations by taking into account interference effects or by

using the extension of ray theory mentioned in §2.1, but at the same time the results correlate well enough to suggest that more detailed field strength measurements would justify fuller analyses. The expected S/N ratio for the direct and the reflected rays was estimated from the approximate relations:

$$S_D = \frac{S_0 \alpha_D}{h_D d_D}; \quad S_R = \frac{S_0 \alpha_R}{h_R d_R}, \qquad \ldots\ldots(31)$$

$$S_{max} = S_D + S_R; \quad S_{min} = |S_D - S_R|, \qquad \ldots\ldots(32)$$

where S = average S/N ratio at range d, between two rays initially inclined to each other at α,

h = height interval between the two rays at range d and initial inclination α,

S_0 = value of S in free space at unit range.

The subscripts D and R stand for the direct and the reflected rays respectively. In equation (31) the earth's divergence factor is neglected and it is assumed that its reflection coefficient is unity. The value of S_0 was obtained by assuming that S_D, S_R are in phase at 30 miles 2000 feet, where the measured signal-to-noise ratio passes through a maximum, with $S_D + S_R = 10$.

The results are shown in figure 8 (*d*). In this figure vertical lines are drawn at a number of ranges and on these lines are marked their points of intersection with the rays. The average values of S_D, S_R, calculated for the height intervals between the intersection points, are written in circles each in its proper interval alongside the height line, and actual values of S_D and S_R are represented by horizontal lines terminated by arrowheads placed against their written values and drawn to the same scale as the representation of the measured values of S/N. Where $S_D + S_R < 1$ the height line is shown dotted. In this way the two quantities, calculated maximum $S/N (= S_D + S_R)$ and measured S/N may be visually compared. The number written at the top of each height line is twice the free space value for S/N at the particular range. The critical rays mentioned in this section are shown by shading. The field at the caustic formed within the shadow region cannot be determined by this method of calculation.

The lower boundary of the shadow, which is also the upper boundary of the radio duct, has been placed some 50–100 feet above the envelope of the rays trapped by the first inversion.

Comparison of the measured and calculated fields shows that

(*a*) The position of the calculated shadow boundaries agrees quite well with the radar results;

(*b*) The measured values of S/N above the duct are, in general, in good accord with the requirement that they lie between $S_D + S_R$ and $|S_D - S_R|$.

Discrepancies which occur at 38 miles and to a lesser extent at 48 miles within the semishadow are larger than the experimental limits and deserve examination. The recorded values of the minimum of S/N occurring at each of these places is 1, whereas the calculated values of $(S_D - S_R)$, although passing through a minimum, only fall to 4 in the first case and 2 in the second. The field is almost entirely due to the reflected ray. A possible explanation of this discrepancy is that it occurs at the commencement of the semishadow where it is narrow, and that edge effects which tend to make the boundaries diffuse may extend to a big proportion of the total thickness. Thus leakage from the

direct rays across both boundaries and partial reflection of the reflected rays at the strong M inversion would both tend to decrease the value of S/N below the lower calculated limit $|S_D - S_R|$.

The methods used cannot of themselves give a quantitative estimate of the complex field distribution within the duct. However, the experimental fact that the signal is a maximum near the maximum of M is in accord with the bunching of rays and the occurrence of caustics at several places near these maxima. A full delineation of the caustics and their cusps would require the drawing of considerably more rays; the intersection of rays belonging to the same pencil may be observed at the following places, points in figure 8 (*c*): direct rays 20 to 35 miles, about 400 feet; 60 to 85 miles, 800–1100 feet; reflected ray, 55 miles (intersection of ray $+6$ and $+6{\cdot}5$).

Although a quantitative estimate of the field distribution within the duct is not feasible by these methods a calculation can be made of the average S/N across the duct at a given range on the assumption that the energy remains trapped without appreciable leakage, but is attenuated through horizontal spreading. It is of interest to note that an average value calculated in this way does lie within the limits observed for S/N at the same range. For instance assuming the angle of the trapped radiation to be 13×10^{-3} radians ($-6{\cdot}5$ to $+6{\cdot}5$) and the vertical spread at 110 miles to be 1300 feet (say from 300 feet to 1600 feet) the theoretical value of S is 2·5; the measured values vary from 4 to 1·5.

§6. CONCLUSIONS

The results of the last section appear to confirm, for frequencies around 200 Mc/s., earlier theories of the existence of shadow zones during certain predictable weather conditions. The results also indicate the utility and practicability of ray tracing methods for obtaining the broad outlines of the field distribution.

It is suggested that those using high-frequency radio systems for the detection or guidance of aircraft in localities subject to strong inversions should keep in mind the possibility, at times, of failure in the system from this cause.

ACKNOWLEDGMENTS

Some of the work described in this paper was carried out as part of the research programme of the Division of Radiophysics, Council for Scientific and Industrial Research, Commonwealth of Australia.

The author wishes to express his great appreciation of the invaluable advice and assistance given by Dr. D. F. Martyn of C.S.I.R. who originated the work. He would like to thank his colleague, Mr. M. Iliffe of the Physics Department, University of Adelaide, who organized the later experimental work, for the many discussions and constructive criticisms which have helped to clarify the problem; and Mr. D. R. Brown, author of O.R.G. (Aust.) Report Number 17, some of whose results are discussed in this paper.

The writer also wishes to thank Dr. E. G. Bowen and Dr. J. L. Pawsey of the Radiophysics Laboratory for making available to him the results of the aircraft soundings and the Laboratory reference and computing facilities.

REFERENCES

Booker, H. G., and Walkinshaw, W., 1946, *Meteorological Factors in Radio-Wave Propagation* (London: Physical Society), p. 80.

Brown, D. R., 1943, *Report* No. 17, *O.R.G.* (Aust.).

ENGLUND, C. R., CRAWFORD, A. B., and MUMFORD, W. W., 1938, *Bell Syst. Tech. J.*, 17, 489.

FREEHAFER, J. E., 1943, *Report* No. 447, *M.I.T. Radiation Laboratories.*

ILIFFE, M. I. G., 1948, in preparation.

MARTYN, D. F., 1943, *Report* No. 19, *O.R.G.* (Aust.).

MARTYN, D. F., and SQUIRES, P., 1944, *Some Factors causing "Superrefraction" on Ultra-High Frequencies in South West Pacific, A.R.P.C. Bulletin.*

PRICE, W. L., 1943, *Report* No. 20, *O.R.G.* (Aust.).

SCHELLENG, J. C., BURROWS, C. R., and FERRELL, E. B., 1933, *Proc. Inst. Radio Engrs.*, 21, 427.

STICKLAND, A. C., 1942, *Dielectric Constant of Water Vapour and its Effect upon the Propagation of very Short Waves*, National Physical Laboratory, Radio Department Report RRB/S2; 1946, *Meteorological Factors in Radio-Wave Propagation* (London: Physical Society), p. 253.

Equivalent Path and Absorption for Oblique Incidence on a Curved Ionospheric Region

BY J. C. JAEGER
University of Tasmania

MS. received 31 *July* 1947

ABSTRACT. The problem of oblique incidence on a curved ionosphere with a Chapman distribution of ionization density is discussed. It is shown that absorption, equivalent path, and the projection of the path on the earth's surface may all be expressed by equivalence theorems of the plane earth type with suitably modified parameters and functions. Numerical values of the functions involved are given.

§1. INTRODUCTION

In a recent paper (Jaeger 1947, referred to below as V) the equivalent path and absorption * for a wireless wave incident vertically on a Chapman region were calculated. This work is extended here to cover oblique incidence on a curved Chapman region in which the ion-density is a function of height only, and the effects of the earth's magnetic field are neglected. Other studies of the oblique incidence problem have either assumed a parabolic distribution of ion-density (Appleton and Beynon 1940, 1947; for other references on the subject see the bibliography in these papers) or have assumed special properties of the region, for example, that it be thin.

The present work consists essentially of an exact solution of the problem of a ray incident obliquely on a curved Chapman region. Since, even for vertical incidence, a final stage of numerical integration is necessary, the accurate integrals for oblique incidence in any particular case are not much more difficult to evaluate than any approximations to them. An approximation is introduced below in order to simplify the presentation of the work by reducing the number of parameters in terms of which the solution is expressed. At all stages the validity of

* Absorption in a Chapman region was first studied by Appleton (1937) who discussed the case of penetration of a region in which μ is effectively unity, and obtained the formula V (26) which I attributed to Best and Ratcliffe (1938)—not (1928) as in V. He also considered the case of V §8 in which ν^2 may be comparable with f^2, and gave the result V (45) with a graph of the integral involved but not the explicit expression V (48) for it.

the approximation has been checked by evaluating the accurate integrals and the discrepancy found to be negligible.

It is found that critical frequency, absorption, equivalent path, and distance traversed along the earth's surface can all be expressed by simple generalizations of the equivalence theorems deduced by Martyn (1935) for a plane ionosphere.

§ 2. FUNDAMENTAL FORMULAE

We have to consider the path of a ray in an ionized region whose ion-density is a function only of h, the height of the point considered above the earth's surface. Let i be the angle of incidence of the ray at height h, and α that at zero height, that is, at the earth's surface (figure 1). Then if R is the radius of the earth, and μ the refractive index at height h, the fundamental equation for refraction in a curved stratified region is

$$\mu(R+h)\sin i = \text{constant} = R\sin\alpha. \qquad \ldots\ldots(1)$$

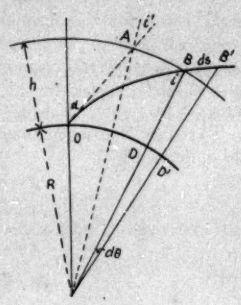


Figure 1.

All the integrals which will be required below involve the quantity

$$\frac{ds}{\mu} = \frac{dh\sec i}{\mu} = \frac{dh}{\sqrt{\{\mu^2 - [R/(R+h)]^2\sin^2\alpha\}}}, \qquad \ldots\ldots(2)$$

where ds is an element of the path of the ray. We proceed to study the quantity (2) in detail; when this has been done the evaluation of the integrals is an easy matter.

The refractive index μ is given by the relation

$$\mu^2 = 1 - N\epsilon^2/\pi m f^2, \qquad \ldots\ldots(3)$$

where f is the frequency of the wave, ϵ and m are the charge and mass of the ions, and N is the ion-density.

In a Chapman region N is given by

$$N = N_0 \exp\tfrac{1}{2}(1 - z - \sec\chi e^{-z}), \qquad \ldots\ldots(4)$$

where χ is the zenith angle of the sun (assumed constant over the region of the ionosphere concerned), and

$$z = (h - h_0)/H \qquad \ldots\ldots(5)$$

is the height above the datum level h_0 measured in terms of the scale height H of the region (cf. V § 2, or Chapman 1931). The quantities N_0, H, and h_0 may be obtained from analysis of vertical incidence data by a method outlined in V. N_0 is the maximum possible value of N, and occurs at height h_c at the equator at noon at the Equinox. The critical frequency at vertical incidence, f_c, was found in V (9) to be

$$f_c^2 = (N_0\epsilon^2/\pi m)\cos^{\frac{1}{2}}\chi. \qquad \ldots\ldots(6)$$

Using (3), (4), and (5) in (2), and making the substitution

$$y=(\tfrac{1}{2}\sec\chi)^{\frac{1}{2}}e^{-\frac{1}{2}z}, \qquad \ldots\ldots(7)$$

so that

$$z=\ln(\tfrac{1}{2}\sec\chi)-2\ln y, \qquad \ldots\ldots(8)$$

(2) becomes

$$\frac{ds}{\mu}=-\frac{2H\,dy}{y}\left\{1-by\exp(-y^2)-\frac{R^2\sin^2\alpha}{[R+h_0+H\ln(\tfrac{1}{2}\sec\chi)-2H\ln y]^2}\right\}^{-\frac{1}{2}} \qquad \ldots\ldots(9)$$

where, as in V(14), we write

$$b=(f_c^2/f^2)\sqrt{(2e)}. \qquad \ldots\ldots(10)$$

To reduce (9) to a more manageable form, we introduce new parameters β, I, d, and a, defined by

$$\beta=1+h_0/R+(H/R)\ln(\tfrac{1}{2}\sec\chi), \qquad \ldots\ldots(11)$$

$$\sin I=(\sin\alpha)/\beta, \qquad \ldots\ldots(12)$$

$$d=b\sec^2 I=(f_c^2/f^2)(2e)^{\frac{1}{2}}\sec^2 I, \qquad \ldots\ldots(13)$$

$$a=(4H/\beta R)\tan^2 I. \qquad \ldots\ldots(14)$$

In terms of these, (9) becomes

$$\frac{ds}{\mu}=-\frac{2H\,dy\sec I}{y\Delta^{\frac{1}{2}}}, \qquad \ldots\ldots(15)$$

where

$$\Delta=1-dy\exp(-y^2)+\tan^2 I\{1-[1-(2H/\beta R)\ln y]^{-2}\}. \qquad \ldots\ldots(16)$$

The expression (15) is still exact; to simplify it further, the orders of magnitude of the parameters involved must be considered. Taking h_0 to be of the order of 120 and 300 km. for E and F regions, respectively, and H to be of the order of 15 and 60 km. in these regions, we find that as χ runs from 0° to 80°, β runs from 1·017 to 1·021 for E region, and from 1·040 to 1·057 for F region. If we allow values of α between 0° and 90°, the maximum values of I given by (12) will be 79·5° and 74° for E and F regions respectively; also I increases steadily to these values as α increases from 0° to 90°.

If $(2H/\beta R)\ln y<1$, we may use the binomial theorem in (16) and obtain

$$\Delta=1-dy\exp(-y^2)-a\ln y\{1+(3H/\beta R)\ln y+\ldots.\}, \qquad \ldots\ldots(17)$$

where a has been defined in (14).

Now in all cases $(H/\beta R)<0{\cdot}01$, so we may neglect all terms of the series in (17) after the first if $0{\cdot}03a(\ln y)^2$ is negligible to our order of approximation. This is the case in all the calculations made here, since in the integrals which have to be evaluated there are additional factors (such as $y\exp(-y^2)$) which are small for the large and small values of y for which $\ln y$ is large and the approximation above might be inadequate.

Thus for purposes of calculation we may replace (17) by

$$\Delta=1-dy\exp(-y^2)-a\ln y. \qquad \ldots\ldots(18)$$

This has the advantage of expressing the whole of the phenomena in terms of the parameters d and a. a is zero for normal incidence or for a plane ionosphere;

it increases from 0 to about 0·27 for E region, and from 0 to about 0·44 for F region, as α increases from 0° to 90°. We shall study the values 0·2 and 0·4 of a which may be regarded as fairly extreme values suited to E and F regions. In § 3 the parameter d will be related to the critical frequency f_α for oblique incidence at α, and the results can thus be expressed in terms of the two parameters f/f_α and a.

§ 3. THE CRITICAL FREQUENCY

For a ray which is transmitted through the ionosphere, $\mu\cos i$ in (2), and thus Δ in (15), never falls to zero. The critical frequency f_α for a ray leaving the earth's surface at α is the greatest value of the frequency f for which Δ vanishes. Thus for a given value of α, and therefore of a, the critical frequency is to be determined from the least value d_a of d for which Δ vanishes. Using the form (18) of Δ, the condition for this is that the equation in y

$$1-a\ln y-yd_a\exp(-y^2)=0, \qquad \ldots\ldots(19)$$

have a double root. This requires that in addition to (19) its derivative shall vanish, that is,

$$a/y+d_a(1-2y^2)\exp(-y^2)=0. \qquad \ldots\ldots(20)$$

From (19) and (20) we find

$$a=\frac{2y^2-1}{(2y^2-1)\ln y+1}, \qquad \ldots\ldots(21)$$

$$d_a=[y\exp(-y^2)\{(2y^2-1)\ln y+1\}]^{-1}. \qquad \ldots\ldots(22)$$

These are a pair of parametric equations for a and d_a in terms of the parameter y.

If $a=0$, $d_a=\sqrt{(2e)}$, and $y=1/\sqrt{2}$, the values for vertical incidence. For oblique incidence only a small range of values of y greater than $1/\sqrt{2}$ is needed to cover the required range $0<a<0{\cdot}5$ of a, and thus the replacing of (16) by (18) is justified. For the relevant range of β and a the deviation of results calculated from (21) and (22) from those from (16) is very small. For example in the rather extreme case $a=0{\cdot}4$, the value of d_a calculated from (21) and (22) is 2·578, while the accurate formula derived from (16) with $a=0{\cdot}4$, and $(4H/\beta R)=0{\cdot}03628$, gives 2·576.

Values of $[\sqrt{(2e)}/d_a]^{\frac{1}{2}}$, calculated from (21) and (22), are shown in figure 2. The critical frequency f_α now follows from d_a by using (13); it is

$$f_\alpha=f_c[\sqrt{(2e)}/d_a]^{\frac{1}{2}}\sec I. \qquad \ldots\ldots(23)$$

For a plane earth this becomes $f_\alpha=f_c\sec\alpha$, the plane earth equivalence theorem; this is seen to generalize immediately to the curved earth on replacing α by I and multiplying by the factor $\sqrt{[\sqrt{(2e)}/d_a]}$ of figure 2 which is nearly unity.

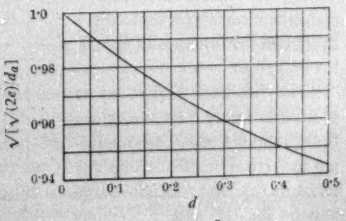


Figure 2.

§4. ABSORPTION IN A DOUBLE PASSAGE OF THE LAYER AT FREQUENCIES GREATER THAN THE CRITICAL FREQUENCY f_α

The total absorption is

$$2\int \kappa\, ds, \qquad \ldots\ldots(24)$$

taken through the layer, where, as in V,

$$\kappa = \frac{\nu}{2\boldsymbol{c}}\frac{1-\mu^2}{\mu} = \frac{N_0\epsilon^2\nu}{2\pi \boldsymbol{c}\mu m f^2}\exp\tfrac{1}{2}(1 - z - \sec\chi e^{-z}), \qquad \ldots\ldots(25)$$

$$\nu = \nu_0 e^{-z}, \qquad \ldots\ldots(26)$$

and $\boldsymbol{c}$ is the velocity of light. It is assumed that $(2\pi f)^2 \gg \nu^2$.

Using (25) and (26) in (24), with the value (15) and (18) of ds/μ, we find

$$2\int \kappa\, ds = H\nu_0 \boldsymbol{c}^{-1}\cos\chi\cos I\ \mathrm{F}(f_\alpha/f;\ a), \qquad \ldots\ldots(27)$$

where

$$\mathrm{F}(f_\alpha/f;\ a) = 4d\int_0^\infty \frac{y^2\exp(-y^2)\,dy}{\sqrt{(1-dy\,\exp(-y^2) - a\ln y)}}. \qquad \ldots\ldots(28)$$

Alternatively, writing

$$\Phi(f_\alpha/f;\ a) = d^{-1}\sqrt{(2e)}\,\mathrm{F}(f_\alpha/f;\ a), \qquad \ldots\ldots(29)$$

(27) may be expressed in the form

$$\begin{aligned} 2\int \kappa\, ds &= H\nu_0\boldsymbol{c}^{-1}\cos\chi\sec I\,(f_c/f)^2\,\Phi(f_\alpha/f;\ a) \\ &= H\nu_0\boldsymbol{c}^{-1}(d_a/\sqrt{(2e)})\cos\chi\cos I\,(f_\alpha/f)^2\,\Phi(f_\alpha/f;\ a). \end{aligned} \qquad \ldots\ldots(30)$$

For vertical incidence $a=0$, and $\mathrm{F}(f_\alpha/f;\ 0)$ and $\Phi(f_\alpha/f;\ 0)$ are the functions $\mathrm{F}(f_c/f)$ and $\Phi(f_c/f)$ defined in (23) and (25) of V and tabulated there.

For a plane ionosphere $a=0$, $I=\alpha$, and (27) becomes

$$2\int \kappa\, ds = H\nu_0\boldsymbol{c}^{-1}\cos\chi\cos\alpha\ \mathrm{F}(f_\alpha/f;\ 0). \qquad \ldots\ldots(31)$$

This is just $\cos\alpha$ times the absorption for vertical incidence at the frequency $f\cos\alpha$ as required by Martyn's equivalence theorem (1935). The modification of this result for a curved ionosphere consists of replacing α by I, and the function $\mathrm{F}(f_\alpha/f;\ 0)$ by $\mathrm{F}(f_\alpha/f;\ a)$.

The function $\mathrm{F}(f_\alpha/f;\ a)$ is plotted in figure 3 for the values 0, 0·2, and 0·4 of a, corresponding to vertical incidence (or plane ionosphere) and rather large values for E and F regions respectively.

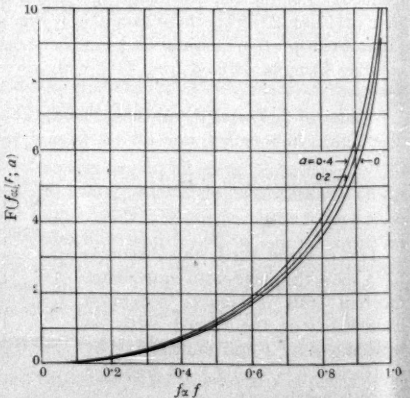


Figure 3. The absorption function for penetration.

§5. ABSORPTION FOR A RAY REFLECTED FROM THE LOWER SIDE OF THE LAYER

In this case we require $2\int \kappa\, ds$ over a range from ground level to the highest point of the path. This point is determined by the value z_1 of z corresponding to the greater root y_1 of

$$1 - dy \exp(-y^2) - a \ln y = 0. \qquad \ldots\ldots(32)$$

The values of y_1 for the values 0, 0·2, and 0·4 of a are shown in the dotted curves of figure 6. Using (5) and (8), the height h_1 of the highest point of the path above the earth's surface is found to be

$$h_1 = h_0 + H \ln(\tfrac{1}{2}\sec\chi) - 2H \ln y_1. \qquad \ldots\ldots(33)$$

Proceeding as in §4, we find for the absorption

$$2\int \kappa\, ds = H\nu_0 \mathbf{c}^{-1} \cos\chi \cos I\, \mathrm{F}_1(f/f_\alpha;\ a), \qquad \ldots\ldots(34)$$

where

$$\mathrm{F}_1(f/f_\alpha;\ a) = 4d \int_{y_1}^{\infty} \frac{y^2 \exp(-y^2)\, dy}{\sqrt{(1 - dy \exp(-y^2) - a \ln y)}}. \qquad \ldots\ldots(35)$$

In figure 4 the function $\mathrm{F}_1(f/f_\alpha;\ a)$ is plotted for the values 0, 0·2, and 0·4 of a. Small values of f are not included, since for these the simple approximation (18) begins to break down. There is no difficulty in extending the curves to small values of f, but there is not much point in doing so because we have also assumed hat $(2\pi f)^2 \gg \nu^2$.

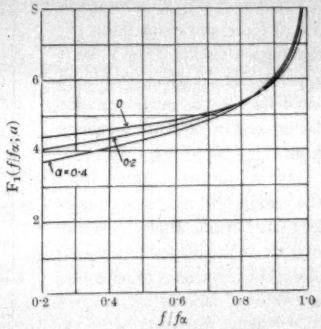


Figure 4. The absorption function for reflection.

It should be remarked that the fact that the curves of figure 4 cross and are rather more widely separated than those of the other figures is caused by the multiplying factor d in the definition (35) of $\mathrm{F}_1(f/f_\alpha;\ a)$. Curves of $[\sqrt{(2e)}/d]\mathrm{F}_1(f/f_\alpha;\ a)$, analogous to Φ of (29), do not cross and lie close together.

§6. THE EQUIVALENT PATH

This is given by

$$\int \frac{ds}{\mu} \qquad \ldots\ldots(36)$$

taken over the whole trajectory of the ray. In V the procedure of evaluating the excess of equivalent path over distance traversed was adopted, and in this way a convergent integral was obtained which was easy to evaluate. A similar simplification can be made for oblique incidence by calculating the excess of the equivalent path over the distance measured to the same height along the tangent to the path at ground level. This distance can be calculated by simple trigonometry.

In figure 1 let OB be the path and OA the tangent to it at ground level. Let i' be the angle this tangent makes with the radius vector at height h, so that

$$(R+h)\sin i' = R\sin\alpha. \qquad \ldots\ldots(37)$$

The excess of equivalent path over distance measured along OA up to any height h is now

$$\int_0^h \left\{ \frac{dh \sec i}{\mu} - dh \sec i' \right\}. \qquad \ldots\ldots(38)$$

Using (15) and (18), and (37) with the approximation leading to (18), this becomes

$$2H \sec I \int_y^\infty \left\{ \frac{1}{(1-a \ln y - dy \exp(-y^2))^{\frac{1}{2}}} - \frac{1}{(1-a \ln y)^{\frac{1}{2}}} \right\} \frac{dy}{y}. \qquad \ldots\ldots(39)$$

For a ray transmitted through the region (which by (1) and (37) emerges with $i = i'$) the excess of equivalent path over distance measured along OA for a double traverse of the region is

$$H \sec I \, \mathrm{P}(f_\alpha/f;\ a), \qquad \ldots\ldots(40)$$

where

$$\mathrm{P}(f_\alpha/f;\ a) = 4\int_0^\infty \left[\frac{1}{(1-a \ln y - dy \exp(-y^2))^{\frac{1}{2}}} - \frac{1}{(1-a \ln y)^{\frac{1}{2}}} \right] \frac{dy}{y}. \qquad \ldots\ldots(41)$$

Curves of $\mathrm{P}(f_\alpha/f;\ a)$ for $a = 0$, 0·2, 0·4 are shown in figure 5; that for $a = 0$ is $\mathrm{P}(f_c/f)$ of V, so it appears that the effect of oblique incidence is to introduce the factor sec I in (40) and to change from $\mathrm{P}(f_\alpha/f;\ 0)$ to $\mathrm{P}(f_\alpha/f;\ a)$.

If the wave is reflected from the region, the excess of equivalent path in the whole trajectory over twice the distance measured along OA to the level of reflection is

$$H \sec I \, \mathrm{P}_1(f/f_\alpha;\ a), \qquad \ldots\ldots(42)$$

where

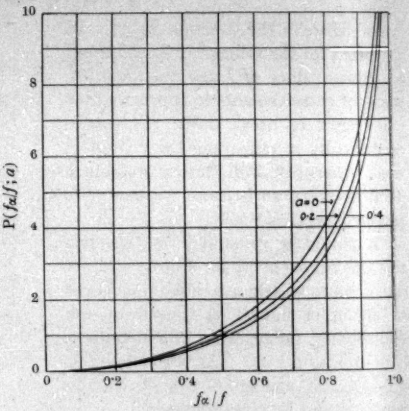


Figure 5. Excess equivalent path for penetration.

$$\mathrm{P}_1(f/f_\alpha;\ a) = 4\int_{y_1}^\infty \left\{ \frac{1}{(1-a \ln y - dy \exp(-y^2))^{\frac{1}{2}}} - \frac{1}{(1-a \ln y)^{\frac{1}{2}}} \right\} \frac{dy}{y}, \qquad \ldots\ldots(43)$$

and y_1 is given by (32). Values of $\mathrm{P}_1(f/f_\alpha;\ a)$ and y_1 are shown in figure 6.

The distance along OA to the level of reflection follows immediately from the trigonometry of figure 1 using the value of h_1 given by (33).

§ 7. THE PROJECTION OF THE PATH OF THE RAY ON THE EARTH'S SURFACE

From figure 1 it follows that the projection DD′ on the earth's surface of the element of path BB′ = ds of the ray is

$$\frac{R\,ds \sin i}{R+h} = \frac{R^2\,dh \sin\alpha \sec i}{\mu\,(R+h)^2}, \qquad \ldots\ldots(44)$$

using (1).

Also, the projection on the earth's surface of the element of OA between the heights h and $h+dh$ is

$$\frac{R\,dh\tan i'}{R+h}=\frac{R^2\,dh\sin\alpha\sec i'}{(R+h)^2}. \qquad \ldots\ldots(45)$$

Thus the horizontal projection on the earth's surface of the ray up to height h is greater than the horizontal projection of the portion of the tangent to the ray at ground level up to height h by the amount

$$\sin\alpha\int_0^h\left\{\frac{\sec i}{\mu}-\sec i'\right\}\frac{R^2\,dh}{(R+h)^2}. \qquad \ldots\ldots(46)$$

As in (39), this is equal to

$$\frac{2H}{\beta^2}\sec I\sin\alpha \times\int_y^\infty\left\{\frac{1}{(1-a\ln y-dy\exp(-y^2))^{\frac{1}{2}}}-\frac{1}{(1-a\ln y)^{\frac{1}{2}}}\right\}\frac{dy}{y[1-(2H/\beta R)\ln y]^2} \qquad \ldots\ldots(47)$$

Thus for a ray transmitted twice through the region, the projection of the path on the earth's surface is greater than the projection of the portion of tangent to the ray at ground level by the amount

$$(H/\beta^2)\sec I\sin\alpha\,Q(f_\alpha/f;\ a;\ 2H/\beta R), \qquad \ldots\ldots(48)$$

where

$$Q\left(\frac{f_\alpha}{f};a;\frac{2H}{\beta R}\right)=4\int_0^\infty\left\{\frac{1}{(1-a\ln y-dy\exp(-y^2))^{\frac{1}{2}}}-\frac{1}{(1-a\ln y)^{\frac{1}{2}}}\right\}\frac{dy}{y[1-(2H/\beta R)\ln y]^2}. \qquad \ldots\ldots(49)$$

The integral Q defined in (49) differs only from P defined in (41) by the factor $[1-(2H/\beta R)\ln y]^{-2}$ in the integrand. Since, for transmission through E region, $2H/\beta R$ is of the order of 0·005 the effect of this factor is small, and it is found that if Q in (48) is replaced by the function $P(f_\alpha/f;\ a)$ given in figure 5, the results obtained are less than 1% too large.

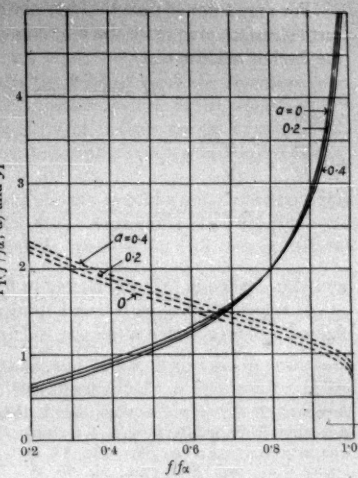


Figure 6.

——— Function $P_1(f/f_\alpha;\ a)$ giving the excess equivalent path for reflection.

– – – – – Function y_1 giving the level of reflection.

For a ray reflected from the layer, the excess of the projection of its path on the earth's surface over twice the projection of the distance measured along OA to the level of reflection is

$$(H/\beta^2)\sec I \sin\alpha\, Q_1(f/f_\alpha;\ a;\ 2H/\beta R), \qquad \ldots\ldots(50)$$

where

$$Q_1\left(\frac{f}{f_\alpha};\ a;\ \frac{2H}{\beta R}\right) = 4\int_{y_1}^{\infty}\left\{\frac{1}{(1-a\ln y-dy\exp(-y^2))^{\frac{1}{2}}}-\frac{1}{(1-a\ln y)^{\frac{1}{2}}}\right\}\frac{dy}{y\sqrt{1-(2H/\beta R)\ln y]^2}}, \qquad \ldots\ldots(51)$$

and y_1 is defined in (32) and plotted in figure 6.

As before, Q_1 may be replaced approximately by the function $P_1(f/f_\alpha;\ a)$ of figure 6; values obtained in this way will be too small, the error being less than 3%.

REFERENCES

APPLETON, E. V., 1937, *Proc. Roy. Soc.* A, **162,** 451.
APPLETON, E. V., and BEYNON, W. J. G., 1940, *Proc. Phys. Soc.*, **52,** 518; 1947, *Ibid.*, **59,** 58.
CHAPMAN, S., 1931, *Proc. Phys. Soc.*, **43,** 26.
JAEGER, J. C., 1947, *Proc. Phys. Soc.*, **59,** 87.
MARTYN, D. F., 1935, *Proc. Phys. Soc.*, **47,** 323.

On a New Type of Rotating Anode X-Ray Tube

BY A. TAYLOR

Research Department, English Electric Co. Ltd.*

* Now Head of the Physics Department, Mond Nickel Co. Ltd., Birmingham

Communicated by Sir Lawrence Bragg; MS. received 11 *December* 1947 *and in amended form* 19 *February* 1948

ABSTRACT. A brief account is given of the physical limitations which influence the power output of stationary and rotating anode x-ray tubes used in crystallographic work. Brief mention is made of various tubes which incorporate rotating vacuum seals in the anode assembly. An account is then given of the "Peristron"†, a continuously water-cooled rotating anode x-ray tube which completely dispenses with the need of rotating vacuum seals. Incorporated in the tube is a novel type of cathode assembly which permits of rapid adjustment to the focus from outside the vacuum envelope.

§1. REQUIREMENTS AND LIMITATIONS ON X-RAY TUBES FOR CRYSTALLOGRAPHIC PURPOSES

In registering x-ray diffraction patterns from crystalline and amorphous materials it is not uncommon to require exposure periods lasting many hours. Several factors play their part in determining the exposure time, only a few of which are under the control of the investigator. To a large extent the exposure time is controlled by the nature of the irradiated specimen and on the type of diffraction pattern required.

† From the Greek "peristrophe" (spinning round like a top) and "electron".

Broadly speaking, there are two direct ways in which the exposure times could be reduced by a considerable amount. The first is to increase the speed of the recording x-ray film, and, if this could be done without introducing undesirable large grain-size effects in the emulsion, it would be an achievement of the highest importance. The second method, which is the subject of this communication, is to increase the intensity of radiant energy at the source. This is the most immediately practical method of obtaining a substantial reduction in exposure period.

With the exception of the Seemann–Bohlin type focusing cameras using curved crystal monochromators or Soller slits, the bulk of x-ray diffraction work is carried out with cameras utilizing pinhole apertures or with powder cameras using narrow slits in their collimator systems. These pinholes or slits must be of sufficient size to bathe the specimen completely in radiation, and, at the same time, confine the beam to prevent unwanted scattering effects. In general, the x-ray beam is admitted into the collimator system through a pinhole aperture, or slit, which seldom exceeds 1 mm. across, and which, in special circumstances, may be as small as 0·25 mm. or less. This means that our first major problem in the design of the x-ray tube is to obtain a focal spot of high intrinsic brilliance and with dimensions comparable with the diameter of the pinhole aperture. Our second problem is to position the focal spot as close to the window of the x-ray tube as possible in order to cut down its distance from the irradiated specimen.

Any increase of focal spot dimensions beyond the size of the pinhole aperture merely represents so much waste of x-ray energy. Thus a tube with a nominal electrical energy consumption of 1 kw. is no better for our special purpose than a tube rated at $\frac{1}{2}$ kw. if only one half of the focal spot is utilized, or if, by the construction of the x-ray tube, the distance of the focal spot from the camera is unduly large. In the final analysis, the ideal solution would be to reduce the size of the focal spot so that it can effectively replace the first pinhole aperture of the camera, the lateral definition of the x-ray beam now being produced entirely by the specimen and the focal spot together.

Of the electrical energy dissipated in the focal spot little more than one thousandth part is converted into useful x-radiation, the remainder merely serves to heat the anode. Thus it is generally unsatisfactory to attempt to produce an intense beam of radiation from a small circular focus, for the local heating soon melts a crater in the anode. In the majority of sealed-off and demountable x-ray tubes of the hot-cathode type, the focus is rectangular in shape and with the approximate dimensions $10 \times 1{\cdot}0$ mm. Such a focus, when viewed along its major axis at a grazing angle of 6° to the anode face, presents a foreshortened area of high intrinsic brilliance approximately 1·0 mm. square.

An x-ray tube with water-cooled copper anode and rectangular focal spot as described above, will normally be operated in the region of 40 kv. with an electron current of 20 ma. This total load of 0·8 kw. may be sustained indefinitely without damage to the anode, although further considerations with regard to the life of the filament may necessitate a somewhat lower tube current. This is specially true in the case of sealed-off tubes.

The energy dissipated in the focal spot is thus in the neighbourhood of 0·08 kw/mm^2. Any effort to increase this loading by concentrating the focal spot into a still smaller area, or by increasing the electrical input, merely raises

the surface of the anode to an unduly high operating temperature. In a tube which has been carefully degassed so that there is no instability from ionization currents, the temperature limit would appear to be determined by the melting point of the anode material. In point of fact, the practical limit is set much lower by the recrystallization of the anode material and the fluctuating thermal stresses to which it is subjected. For the case of copper, de Graaf and Oosterkamp (1938) give an upper limit of 300° c.

It was realized very early in the history of x-ray tubes that a substantially larger output of radiation would be possible if the anode were set in motion thereby continuously exposing a fresh, cool region to the electron bombardment (Breton 1897). This has been achieved by oscillating the target with a rotary movement of a few degrees (Hosemann 1939) and by a slow translatory oscillation of several centimetres (Astbury and MacArthur 1945). It has also been achieved by a gyratory oscillation which gives the effect of continuous rotation (Du Mond and Youtz 1937) and by a true continuous rotation of the anode (Coolidge 1915, 1930, Elihu Thomson 1914, Bouwers 1929). Most ingenious of all is the method in which the entire x-ray tube rotates about the cathode–anode axis while the electron stream, deflected by an external magnet, strikes the anode at a fixed position in space (Breton 1897, Coolidge 1917). Few of these arrangements, are, however, satisfactory for x-ray crystal analysis work.

Müller (1927, 1929, 1931, 1939) has made a detailed mathematical analysis of the heat dissipation in water-cooled stationary and rotating x-ray anodes. Within certain limits, he has shown that for a line focus

$$W_{\max}\ (\text{rest}) = 4{\cdot}25(T-T_0)k\lambda(a,b), \qquad \ldots\ldots(1)$$

$$W_{\max}\ (\text{moving}) = \frac{4{\cdot}04(T-T_0)ka\sqrt{(vb\rho c/k)}}{1+(2a/\pi^2R)(1+\ln 4R/a)\sqrt{(vb\rho c/k)}}, \qquad \ldots\ldots(2)$$

$$\simeq 4{\cdot}04(T-T_0)ka\sqrt{(vb\rho c/k)}, \qquad \ldots\ldots(3)$$

where $W_{\max}$ = limiting input in watts, T = maximum safe temperature of surface T_0 = temperature of cooling metal-water interface, a, b = half length (width) of focal spot—assumed elliptical, k = thermal conductivity of target material (watts/cm/°c.), ρ = density of target material, c = specific heat of target material (watt.sec/gm/°c.), v = relative velocity of focal spot (cm/sec.), R = radius of path of focal spot on target, $\lambda(a,b) = \lim_{n\to\infty} A_n = \lim_{n\to\infty} B_n$, where $A_n = \frac{1}{2}(A_{n-1}+B_{n-1})$ and $A_0 = a$, $B_0 = b$.

Thus, by giving the anode the velocity v relative to the focal spot, we can increase the performance of the x-ray tube in the ratio

$$\frac{W_{\max}\ (\text{moving})}{W_{\max}\ (\text{rest})} = \frac{0{\cdot}951a\sqrt{(vb\rho c/k)}}{\lambda(a,b)\{1+(2a/\pi^2R)(1+\ln 4R/a)\sqrt{(vb\rho c/k)}\}}. \qquad \ldots\ldots(4)$$

In the case of a copper anode, $\rho c/k \simeq 1$. For a focus of dimensions $2a = 0{\cdot}8$ cm, $2b = 0{\cdot}05$ cm, $\lambda(a,b) = 0{\cdot}15$. With an anode of radius $R = 9{\cdot}0$ cm. which rotates at 1370 r.p.m., we easily find that $W_{\max}$ (moving)/$W_{\max}$ (rest) = 11·3. Equation (4) from which the value 11·3 is derived assumes, among many things, that the x-ray tube is operated at a steady potential. It does not follow then, that if A.C. is applied across the tube the same increase in efficiency is to be expected: in practice, the output is raised by a factor more nearly approaching half the theoretical limit.

§2. TYPES OF ROTATING ANODE TUBE

In the case of water-cooled rotating anode tubes we are faced with the dual problem of rotating the anode within the vacuum envelope, and, at the same time, introducing a continuous flow of cooling water into the anode. A number of solutions to this problem have been described in the literature.

The first water-cooled rotating anode tube, developed by Stintzing (1926), was designed primarily for X-ray spectroscopy and was not suitable for crystallographic work. The X-ray tubes developed by Müller and Clay (Müller 1929 b, Clay 1934, Müller and Clay 1939) use a cooling system closely resembling that of Stintzing. The anode is in the form of a flat hollow cylinder driven by a hollow axle which is coupled to an electric motor and rotated at 2000 r.p.m. The flat face of the cylinder, on the side opposite the driving axle, serves as the effective anode surface, the focus marking out a continuous narrow track near its periphery. Cooling water is injected through a co-axial inlet tube housed within the hollow driving shaft and is forced against the inside wall of the anode periphery and then returns by way of the hollow shaft to the outlet pipe. Because of its fundamental importance the arrangement in Clay's 5 kw. X-ray generator is shown diagrammatically in figure 1.

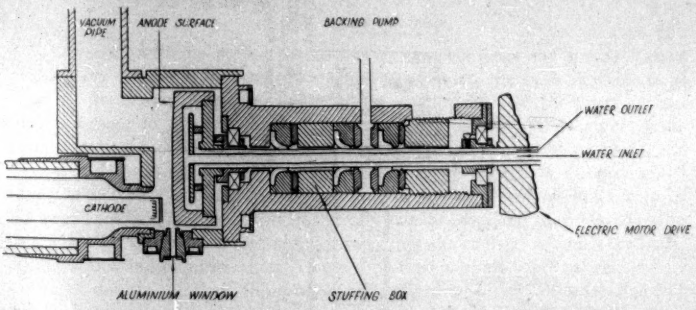


Figure 1. Rotating anode X-ray tube with stuffing box vacuum seal. (Clay.)

Perhaps the most important part of the assembly is the vacuum seal surrounding the rotating shaft. This consists of a stuffing-box packed with hat-leathers pre-treated with vacuum grease and backed with high-speed vacuum pumps. A very similar arrangement is described by Beck (1939).

A stuffing-box type of seal has a number of drawbacks. It leaks when the backing pumps are shut down, and the frictional forces, which produce an appreciable wear on the driving shaft, must be overcome by employing a 2 H.P. electric motor.

These drawbacks led Astbury and Preston (1934) to make use of an almost frictionless permanent liquid seal. This takes the form of a column of mercury of barometric height surrounding the hollow driving shaft of the anode. A detailed description of the apparatus and its performance is given by MacArthur (1944).

A satisfactory and compact seal is provided by a greased conical vacuum joint on the rotating hollow anode shaft. (Fournier, Gondet and Mathieu

1937). The speed of rotation is relatively small (60 r.p.m.), and this limits the energy output.

A description is given by Linnitzki and Gorski (1936) of two rotating anode tubes which eliminate the conventional rotating seals on the driving shafts, replacing them with molecular pumps of the combined Gaede–Hollweg or Gaede–Siegbahn type. No performance data for these tubes are given.

From the brief descriptions of the rotating anode x-ray tubes given above, one feature common to them all becomes evident. That is, the cooling water

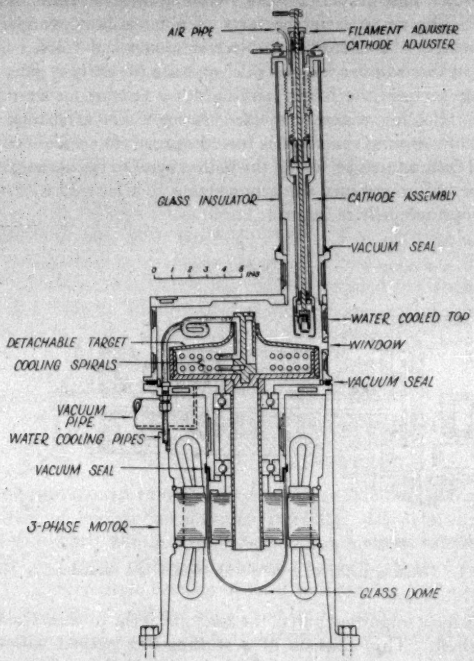


Figure 2. Section through x-ray tube.

is led to the inner surface of the anode through a hollow driving shaft, which, in its turn, must pass the wall of the x-ray tube by way of some form of rotating vacuum joint. The Peristron (Taylor 1944) described below, is a rotating anode x-ray tube using continuous water cooling, and eliminating entirely the need for rotating vacuum seals.

§ 3. DESCRIPTION OF THE PERISTRON

Figure 2 is a vertical section through the rotating anode x-ray tube. Motive power to the rotating anode is given by a 3-phase 400-volt induction motor having a specially large air-gap 5/32″ across and therefore consuming only a

small fraction of 1 H.P. The speed of rotation is 1370 r.p.m. The brass base of the vacuum envelope of the X-ray tube is accurately spigotted into the cylindrical top of the induction motor which serves as a very rigid support. Brazed into the base is a right-angled vacuum pipe $1\frac{1}{2}$" in diameter. The centre of the base houses the ball-races for the driving shaft in a central tubular member, the lower end of which is terminated by a waxed-on $\frac{1}{8}$" thick glass dome. This occupies the air-gap of the motor and effectively seals the rotor, at the lower end of the driving shaft, within the vacuum envelope.

Screwed into the upper end of the shaft is the detachable hollow anode. The whole sub-assembly, comprising anode, base-plate, rotor and glass dome, removed from the motor housing, is shown in figure 3.

The water-cooled brass top of the X-ray tube is vacuum sealed to the base-plate by means of a flat neoprene gasket. The flanges gripping the gasket in position are turned with a series of fine concentric ridges. These bite into the neoprene so that the atmospheric pressure bearing down on the top is quite sufficient to maintain a perfect vacuum seal which is both easily demountable and trouble-free. The cathode assembly is mounted on the top by means of a demountable plasticine vacuum joint. The window, which is of lithium metal 1 mm. thick, is pressed into position by a screw cap and is easily replaced.

(i) *The anode assembly*

The anode with its associated cooling system comprise the special feature of the X-ray tube. It is essentially a hollow cylinder, the base of which is of brass $8\frac{1}{2}$" in diameter and $\frac{1}{4}$" thick, with a screwed boss on the lower surface for attachment to the driving shaft of the electric motor. The upper anode surface is made of pure copper sheet, $\frac{1}{8}$" thick and with a slope of 6°, the focal spot delineating a track approximately 10 mm. wide at a distance of $\frac{1}{4}$" from the periphery. In the original construction, the bush at the centre of the anode surface and the cylindrical walls were of brass and were brazed to the copper anode, but a much simpler construction was evolved in which the bush, anode surface and cylindrical walls were spun from a single copper sheet.

A flat double-spiral of thin-walled copper tubing, 5/16" external diameter and $\frac{1}{4}$" bore is held stationary within the hollow anode by the spindle which passes through the centre bush in the upper surface of the rotating anode. Water is introduced through pipes passing through stationary demountable seals in the base of the vacuum envelope and passes, in turn, through the spindle into the spirals within the anode and out again through the spindle.

The water-cooling leads to the spirals are bridged together by a brass plate which bears on a rubber gasket, thus forming a demountable vacuum seal covering the holes where the nipples on the pipes pass through the base. The bridge also serves to prevent any twisting of the pipes as the locknuts on the vacuum seals and pipe nipples are tightened.

The anode is partly filled with vacuum oil through the large inlet hole drilled in the spindle which is then closed by means of a screwed filler-cap. About $\frac{1}{2}$ litre of oil is sufficient. The surface of the oil and its vapour are in active communication with the high vacuum within the X-ray tube through the gap between the spindle and the bushing. Thus it is necessary to use an oil with an extremely low vapour pressure such as Apiezon "B", Octoil "S" or a silicone. In the present instance Apiezon "B" with its vapour pressure of 10^{-6} mm.

has been successfully used. Any possible creep of oil through the gap between the spindle and bush is prevented by a screw thread cut into the bush which drives the oil back into the anode. A clearance of about 10/1000″ is required to prevent the bush seizing on the spindle, since the vacuum oil has very poor lubricating properties.

In the course of rotation, the centrifugal forces fling the oil against the circumferential inner surface of the anode. By this means, the heat generated in the metal surface of the anode by the electron bombardment is instantly transferred to the oil, which in its turn is transferred to the cooling water flowing through the stationary spirals. Moreover, the churning motion imparted to the oil as it is forced past the spirals provides an extremely efficient method of transferring the heat, for the cooling is not dependent on natural convection currents as in a static body of liquid. With a flow of water of only 10 litres per minute, and an initial temperature of 14° c., the outflow temperature was 20° c. when the power dissipated in the anode was 5 kw.

The design of the x-ray tube is such that the anode is easily accessible for cleaning and for replacement whenever different characteristic radiation is required. Upon lifting away the lid from the rubber gasket which forms the vacuum seal with the base-plate, the whole anode assembly is exposed (figure 4). To remove the anode, the water-cooling pipe nipples and locknuts on the base-plate are unscrewed and then the anode is unscrewed from the driving shaft, which is held stationary by means of a flat spanner. The driving shaft, bearings and rotor are thus left behind as an integral part of the base-plate and need never be disturbed. The time required to strip down the tube and replace the anode is little more than 5 minutes.

After re-assembly it is important to rotate the anode before commencing pumping operations. This prevents any tendency for oil to creep past the bush should it foam during the escape of adsorbed air. For this reason a thermal delay switch was incorporated in the electrical controls which gave a few seconds for the anode to rotate up to speed before allowing the rotary pump to start. Once vacuum conditions were established the anode could be rotated or stopped at will independently of the pumps.

The x-ray tube is mounted on a steel stand which houses the vacuum pumps and also serves to carry an electrical control panel. The pumping unit comprises Metropolitan Vickers O2 and O3 oil condensation pumps mounted in cascade and backed by a DR 1 type rotary pump. The cooling water from the pumps is arranged to pass into the cooling spirals of the anode. This is a more efficient arrangement than passing the water in the reverse direction since the temperature rise on passing through the pumps is negligibly small. It has the advantage over duplicate sets of cooling leads that only one water relay safety switch is required in the electrical circuit.

Directly beneath the platform supporting the x-ray tube are two small air compressors. One circulates a stream of cooling air round the stator windings and glass dome of the x-ray tube, the other serves to cool the cathode assembly.

(ii) *The cathode assembly*

In the majority of demountable x-ray tubes, the provision made for adjusting the position of filament and focusing tube with respect to the anode is very crude, and any alterations must be made by opening up the vacuum system. This is a time-consuming process with the result that an operator is too often content

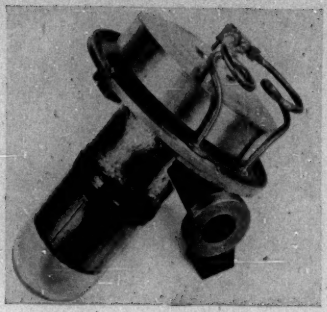


Figure 3. Anode sub-assembly removed from motor-housing.

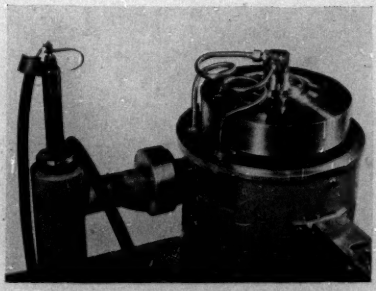
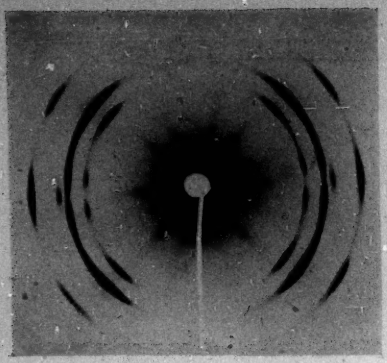
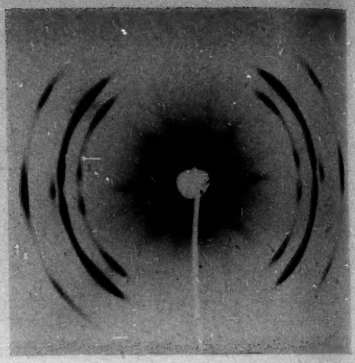


Figure 4. Rotating anode x-ray tube with top removed, exposing anode assembly.

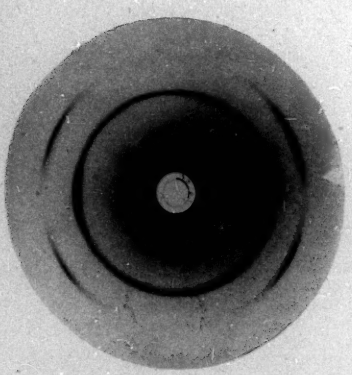
To face page 92



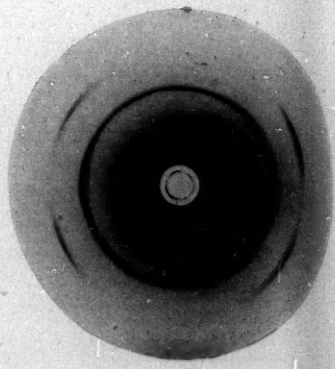
(*a*)



(*c*)



(*b*)



(*d*)

Figure 6. Transmission (*a*) and back-reflection (*b*) photographs of 0·0015" copper foil, taken with stationary anode x-ray tube. CuK$\alpha+\beta$ radiation. 40 kvp. 20 ma. Exposure time 20 minutes. (*c*) & (*d*) Rotating anode tube. 40 kvp. 100 ma. Exposure time 3 minutes. (Reproduced $\frac{1}{2}$ size.)

to let the tube run with an inefficient broad focus. Moreover, it is not generally realized that if the tube is focused for running with a chromium anode at 35 kv. a change to 90 kv. for operating with a silver or tungsten anode will have a marked defocusing effect. It was in order to overcome these difficulties that a new adjustable type of cathode assembly was developed.

In the present arrangement shown in figure 5, the position of the cathode assembly relative to the anode is altered by means of a flexible metal bellows moved by a nut, the cathode adjuster, situated outside the vacuum system on the metal cap sealed to the glass insulator. A second internal bellows is connected to the filament holder, which may be moved up or down within the focusing tube by means of the filament adjuster nut, which is soldered to a porcelain insulator above the cathode adjuster. It is therefore possible to make adjustments and obtain the best focusing conditions in a few seconds by observing the changes in a pinhole image of the focal spot upon a fluorescent screen—an operation which normally takes several hours with the more orthodox type of cathode arrangement.

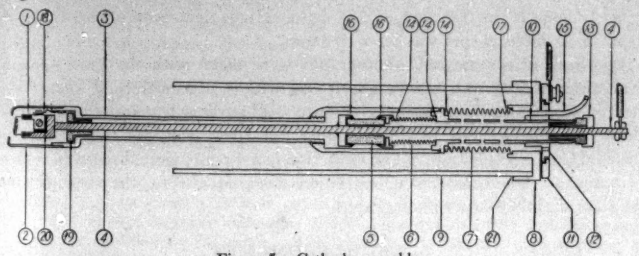


Figure 5. Cathode assembly.

List of Parts

1. Filament assembly
2. Focusing tube
3. Steel tube
4. Filament lead
5. Porcelain insulator
6. Flexible bellows
7. Flexible bellows
8. Brass cap
9. Glass insulator
10. Screwed tube
11. Clamping ring
12. Cathode assembly adjusting nut
13. Filament retractor nut
14. Soldered vacuum joint
15. Air pipe
16. Metal-porcelain vacuum joint
17. Bellows-cap vacuum joint
18. Flexible filament lead
19. Porcelain guide
20. Mica insulation
21. Wax vacuum seal

Experiments were made with a number of filament sizes and types of focusing tubes. Ultimately a simple combination was found which gave a sharp focus of 10 mm. × 0·5 mm. at 45 kv., and 100 ma. This consisted of a tungsten spiral of 16 turns of 0·33 mm. diameter wire wound on a mandril 0·8 mm. diameter into a helix 10 mm. long, housed within a steel focusing tube with a simple slot 2½ mm. across and 12 mm. long. The slot opens out from the surface nearest the anode towards the filament at an angle of 45°. The focusing tube stands 15 mm. from the target with the filament surface approximately 1 mm. behind the front surface of the slot. This latter distance is most critical, for a movement of $\frac{1}{10}$ mm. in either direction will produce considerable changes in the sharpness of focus. With the target bevelled at 6° the effective focal spot is a rectangle 1 mm. high by 0·5 mm. across. A filament current of 8·5 amp. is sufficient for a tube current of 100 ma. Excellent results have also been obtained with tantalum filaments.

The filament is held in position in the simplest possible manner. The straight ends of the filament spiral are pushed into 1 mm. diameter holes drilled

in the filament holder and held rigidly in position by driving in thin soft iron wedges. These have the great advantage of being easily removed when the filament is due for replacement. Since an appreciable amount of heat is conducted from the filament along the steel cathode assembly to the soldered joints on the bellows and to the Apiezon "W" wax seal between the glass insulator and the end cap, some form of cooling is desirable. A stream of cooling air, generated by a small compressor mounted below the x-ray tube, is circulated through the bellows system via an inlet pipe soldered into the cathode adjuster nut. Such cooling only becomes necessary if the tube is to be run continuously for periods of over an hour.

§4. PERFORMANCE OF THE X-RAY TUBE

The x-ray tube may be operated continuously at 100–120 ma. and 45 kv. without giving any trouble. It has also been run for varying periods at 30 kv. at 175 to 200 ma. It must be stated that these figures are nominal in that raw A.C. is applied from the H.T. transformer directly to the tube and the resultant D.C. component in the transformer may cause serious distortion of the wave form. In this work the focal spot was $0{\cdot}5 \times 10{\cdot}0$ mm.

As a basis of comparison, photographs were taken with the same specimen in the same camera with the Peristron and with a new sealed-off Victor x-ray tube run at its maximum output. Figure 6 illustrates transmission and back reflection photographs of $1\frac{1}{2}/1000''$ copper foil taken with the two tubes using unfiltered Cu $\alpha + \beta$ radiation. The reduction in exposure time is from 20 minutes to 3 minutes. On account of its extremely sharp focal spot, the rotating anode tube gives slightly sharper photographs.

ACKNOWLEDGMENTS

The author wishes to thank the English Electric Co. Ltd., in whose research laboratories the Peristron was developed, for their kind permission to publish this paper, and Mr. J. K. Brown, Director of Research, for his very helpful interest. He also wishes to acknowledge the great assistance given by Mr. F. Leighton in the construction and operation of the equipment.

REFERENCES

ASTBURY, W. T., and MACARTHUR, I., 1945, *Nature, Lond.*, **155**, 108.
ASTBURY, W. T., and PRESTON, R. D., 1934, *Nature, Lond.*, **133**, 460.
BECK, J., 1939, *Phys. Z.*, **40**, 474.
BOUWERS, A., 1929, *Verh. dtsch. Röntgen-Ges.*, **20**, 102.
BRETON, J. L., 1897, Les Rayons Cathodiques et Rayons X. *Rev. Sci. et Ind.*, p. 86.
CLAY, R. E., 1934, *Proc. Phys. Soc.*, **46**, 703.
COOLIDGE, W. D., 1915, *Amer. J. Roentgenol.*, 884; 1917, U.S. Patent No. 1215116; 1930, *Gen. Elect. Rev.*, **33**, 608.
DU MOND, J. W., and YOUTZ, J. P., 1937, *Rev. Sci. Instrum.*, **8**, 291
FOURNIER, CONDET and MATHIEU, 1937, *J. Phys. Radium*, **8**, 160.
DE GRAAF, J. E., and OOSTERKAMP, W. J., 1938, *J. Sci. Instrum.*, **15**, 293.
HOSEMANN, R., 1939, *Z. tech. Phys.*, **20**, 203.
LINNITZKI, V., and GORSKI, V., 1936, *Tech. Phys. U.S.S.R.*, **3**, 220.
MACARTHUR, I., 1944, *Electronic Engng.*, **17**.
MÜLLER, A., 1927, *Proc. Roy. Soc.* A, **117**, 30; 1929 a, *Ibid.*, **125**, 507; 1929 b, *Nature, Lond.*, **124**, 128; 1931, *Proc. Roy. Soc.* A, **132**, 646.
MÜLLER, A., and CLAY, R. E., 1939, *J. Instn. Elect. Engrs.*, **84**, 261.
STINTZING, H., 1926, *Phys. Z.*, **27**, 844.
TAYLOR, A., 1944, British Patent Applications 25718/44 and 37511/46.
THOMSON, ELIHU, 1914, U.S. Patent, described in Coolidge (1930).

LETTERS TO THE EDITOR

The Dipole Moment of the Supposed Fundamental Magnetic Field due to Rotation

In his recent important discussion of the magnetic fields of the earth, sun and stars, Prof. Blackett (1947) suggested that these fields represent a new and fundamental property of rotating matter. His formulation of his hypothesis (reviving and revising one already proposed by H. A. Wilson) is at present tentative, and applicable only to bodies symmetrical about their axis of rotation; it is that ρv, the mass flux due to the rotation, is associated with a fundamental magnetic field as if it were an electric current intensity $K\rho v$; here ρ denotes the mass density, v the rotating velocity ωp (ω the angular velocity and p the distance from the axis), and K, when the current intensity is expressed in electromagnetic units, is given as $\beta G^{\frac{1}{2}}/2c$ (G the constant of gravitation, c the speed of light, and β a factor of order unity).

As Blackett indicated, the external fundamental field of a *spherical* body in *uniform* rotation, when the density distribution is *spherically symmetrical*, is of dipole character, and the following relation holds between the dipole moment P and the angular momentum U: $P/U=\frac{1}{2}K$. Thus P/U is independent of the nature of the function $\rho(r)$ expressing the spherically symmetrical density distribution.

In this note I show, very simply, that $P/U=\frac{1}{2}K$ for a rotating body when there is symmetry about the axis of rotation, whatever the form of the body, or the non-uniformity either of ρ or ω.

Let the position of any point of a half meridian section (bounded by the axis) be specified by p, the distance from the axis, and z, the coordinate measured parallel to the axis. Let $dp\,dz$ be the area of an element of the section, with sides parallel and perpendicular to the axis, at the point p, z. The equivalent electric current associated with this element is $K\rho\omega p\,dp\,dz$; it flows round a circle of radius p. By Ampère's hypothesis its field is the same as that of the magnetic induction B for a uniform magnetic shell of strength $j=K\rho\omega p\,dp\,dz$ bounded by the circle; it is convenient to take this shell as plane; its magnetic moment is j times its area (πp^2) and is parallel to the axis. As moment vectors are additive, and as these moments are all in the same direction along the axis (except in so far as ω may vary in sign), the total dipole moment P is given by

$$P=\Sigma\pi p^2 j=\int\int\pi K\rho\omega p^3\,dp\,dz.$$

In this expression ρ and ω may be any functions of p and z.

The mass of the ring associated with the element $dp\,dz$ is $2\pi p\rho\,dp\,dz$, and its speed is ωp; its angular momentum is therefore $2\pi\rho\omega p^3\,dp\,dz$, and thus by integration $U=2\pi\int\rho\omega p^3 dp\,dz$. Hence $P/U=\frac{1}{2}K$, as stated.

If C is the centre of the current ring associated with the element $dp\,dz$ (so that C is the axial point with coordinate z), the potential of the field of the ring, at a distant point, can be expressed as a series of zonal harmonic terms, of which the first is determined by the dipole moment of the ring. If the potential is expressed relative to any other point O on the axis, this first term is unaltered, but the coefficients of all the higher harmonics are changed. When the contributions of all the rings are combined to give the whole potential, the first harmonic corresponds to the dipole moment $\frac{1}{2}KU$, and in addition there are in general higher harmonics which do not vanish whatever the choice of O. These do vanish, however, for a spherical body rotating about a diameter, when ρ and ω are functions only of the radial distance.

In a paper on "Solar magnetism and the hypothesis of fundamental magnetization by rotation", to be communicated to the Royal Astronomical Society, some examples of these higher harmonics are given for a non-uniformly rotating sun.

Queen's College, Oxford.
26th April 1948.

S. CHAPMAN.

BLACKETT, P. M. S., 1947, *Nature, Lond.*, **159**, 658.

The Freezing-in of Nuclear Equilibrium

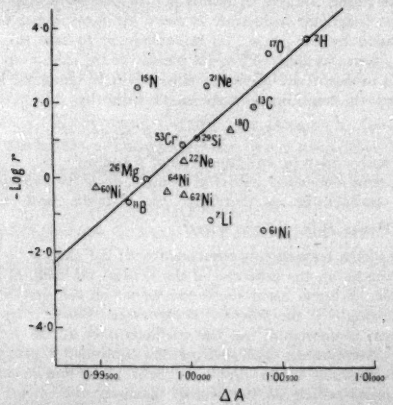
In a previous paper (Ubbelohde 1947) a method was described for testing whether the abundance ratios of stable isotopes might be due to the freezing-in of thermodynamic equilibrium. The test was constructed so as to avoid uncertainties about the density of the system, and about the mass of the neutron, at the time of freezing-in. From the freezing-in curve a freezing-in temperature was calculated of approximately 10^{10} °K.

Further data have since become available (Mattauch 1946, Stephens 1947). Where the masses of the stable isotopes are known to at least four places of decimals, it is possible to test whether the additional abundance ratios are in agreement with the freezing-in curve. Uncertainties in ΔA greater than 0·0002 prevent any decision for all the elements with $Z>28$. No conclusions can yet be reached about their non-thermodynamic origin, on account of this gap in the data.

New values of ΔA of sufficient accuracy for the test include the following :

$^{m+1}A_n$	Abundance ratio $(m+1)/m$	ΔA (mass units)	$^{m+2}A_n$	Abundance ratio $(m+2)/m$	ΔA (mass units)
^{3}He	10^{7}	0·986872	^{18}O	0·0020	2·00485
^{11}B	4·00	0·996732	^{60}Ni	0·396	1·99010
^{13}C	0·011	1·003681	^{62}Ni	0·142	1·99978
^{15}N	0·00384	0·997340	^{64}Ni	0·221	1·99785
^{17}O	0·0004	1·00450			
^{61}Ni	0·0449	1·00419			

When these values are plotted on the curve of $-\log r$ against ΔA, previously established (Ubbelohde 1947) (see figure), it will be seen that for the isotopes ^{11}B, ^{13}C, ^{17}O, ^{18}O, ^{60}Ni, ^{62}Ni, ^{64}Ni there is good agreement within the limits of uncertainty previously found. (To avoid confusion, not all the earlier data are included in the figure.) For the isotopes ^{3}He, ^{15}N and ^{61}Ni no thermodynamic relation can be found, unless the reactions maintaining equilibrium are very notably slower and faster respectively than the average.



Activity of the neutron at freezing-in

The value of the intercept made by the freezing-in curve when $\log r=0$ is $\Delta A=0{\cdot}99780$. This intercept may be used in conjunction with the freezing-in temperature and an estimated value for the mass of the neutron, to calculate the neutron activity at the time of freezing-in.

For convenience, the equilibrium equations of the previous paper are rewritten :

$$-\log n+\log r=-4{\cdot}7\times10^{13}(\Delta A-m_n)/T+3/4{\cdot}606-\tfrac{3}{2}\log T-\log[(2\pi m_n k)^{\frac{1}{2}}k/h^3], \quad \ldots\ldots (1)$$

where the effects of the difference in mass between isotope pairs, and any differences in spin multiplicities have been neglected in the last term.

Writing $\log r=0$, $T=10^{10}$, $\Delta A=0{\cdot}99780$, $m_n=1{\cdot}008941$ (cf. Stephens 1947), we find $-\log n=-13{\cdot}87$. There seems to be no likelihood of a value of n as large as this. The numerical result is, however, very sensitive to the exact numerical value taken for the freezing-in temperature. It is rather less sensitive to the value assumed for the mass of the neutron. For example, if a freezing-in temperature of 10^9 were used, $-\log n=+34{\cdot}9$.

This value of n is very much too small to be probable. Clearly a temperature intermediate between 10^{10} °K. and 10^9 °K. would give quite plausible values of n. But the test for the thermodynamic equilibrium is best constructed so as to avoid the use of n altogether, as indicated in the previous paper.

It is important to note that the very rapid change in neutron concentration over the temperature range between 10^{10} °K. and 10^9 °K. is imposed by the physical magnitudes in equation (1). This rapid change provides a simple explanation for the suggestion made in the previous paper (Ubbelohde 1947) that freezing-in occurred on account not of a deficiency in activation energy, but of a deficiency in neutron concentration at around the freezing-in temperature.

With regard to the effect of the mass of the neutron, using the values given above $\Delta A-m_n=-0{\cdot}01114$. The estimated uncertainty in m_n is $\pm 0{\cdot}00002$, which would not affect the calculations appreciably. The maximum uncertainty in the ΔA intercept is about $\pm 0{\cdot}00150$, and this would not be enough to make a serious difference in n. But at the temperature of 10^{10} °K., the R.M.S. velocity of particles of unit mass ($^{16}O=16$) is approximately $1{\cdot}5\times 10^9$ cm/sec., i.e. $0{\cdot}05\beta$. Neutrons hotter than the average will begin to exhibit relativity effects, which may to some extent modify equation (1).

Thanks are due to Mr. G. E. Roberts and Mr. J. K. Brown for drawing my attention to the additional data quoted above.

Queen's University, Belfast. A. R. Ubbelohde.
25th March 1948.

Mattauch, J., 1946, *Nuclear Physics Tables* (Interscience Publ. Inc.).
Stephens, W. E., 1947, *Rev. Mod. Phys.*, **19**, 19.
Ubbelohde, A. R., 1947, *Proc. Phys. Soc.*, **59**, 139.

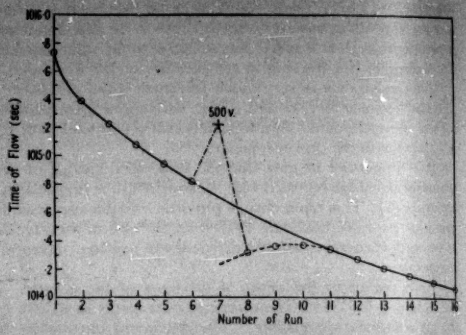
The Detection of Heating in Liquid Dielectrics at Low Frequencies

The heating in dielectrics which occurs at radio frequencies is well known and is so marked that its detection presents no experimental difficulty. It has indeed found many applications in industry since the uniform production of heat throughout the substance is difficult, if not impossible, to produce by other methods. The heating is very dependent on the frequency of the alternating electric field, and at low frequencies, say 10^4 c/sec., becomes so small that its detection and accurate measurement are extremely difficult.

During the course of some experiments by Prof. E. N. da C. Andrade and myself on the effect of alternating electric fields on the viscosity of certain liquids, the dielectric heating of certain polar liquids not only showed itself clearly but proved to be capable of rough quantitative estimation at frequencies as low as 10^4 c/sec. In these experiments a viscometer is used in which the liquid flows through a narrow rectangular channel between massive steel plates (Andrade and Dodd 1946). The viscometer is contained in an air bath immersed in a large water bath whose temperature is controlled so that its variation is less than 1/1000 degree over a period of many hours. The time of flow of a certain volume of the liquid through the channel is measured by the successive interruptions of two beams of light by the liquid falling in one limb of the viscometer. The amplified electric impulses from the photoelectric cells upon which the light beams are directed are applied to high-speed relays used to start and stop a synchronous electric clock reading to 0·001 second. Times of flow taken at four-hourly intervals do not differ by more than 0·01 second for a total time of flow of over 1000 seconds, showing that evaporation is negligible.

Alternating electric fields of amplitude 50 kv/cm. and frequencies up to 20,000 c/sec. have been applied to the liquid whilst it was flowing through the gap, the times of flow of the liquid being measured not only while the field is applied but also both before and after the application of the field.

With monochlorbenzene, times of flow taken immediately after one another are plotted in the figure, the field of amplitude 35 kv/cm., and frequency 10^4 c/sec., being applied after six measurements without field. The steady decrease of the first six values has been shown to be due to the slight heating of the liquid caused by the dissipation of the work done in forcing it through the capillary. A much bigger decrease occurs immediately after the application of the field, indicating that a new form of heating has occurred. This heat is slowly absorbed by the thermostat, and after four or five runs the times fall accurately on the curve through the points obtained before the application of the field. This can only be dielectric heating as the liquid is an almost perfect insulator, so that the ordinary Joule heating is absent. Confirmation of this arises from the fact that no such decrease in time of flow was observed immediately after the application of a *steady* field of 40 kv/cm. The increase in the time of flow which occurs whilst the field is being applied is, of course, due to the increase in the viscosity of the liquid and will be fully discussed elsewhere.



The dielectric heating which occurs whilst the field is being applied can be obtained by extrapolation and is seen to cause 0·46 sec. decrease in the time of flow of 1015 sec. With the liquid in question this represents a mean temperature rise of about 0·04°.

The purpose of this note is to draw attention to this method of detecting very small dielectric heating in liquids. The thermal capacity and surface area of the electrodes, although appropriate for the main purpose for which they were designed, are too large for the exact estimation of the dielectric heating. However, a rough estimation can be made of the power factor at these low frequencies on the following considerations.

The heat generated in the liquid will be communicated to the steel plates forming the viscometer. The velocity of propagation of a simple harmonic temperature wave of period T through the metal of diffusivity h is $\sqrt{(4\pi h/T)}$. Taking h for steel as 0·15 $cm^2\ sec^{-1}$ and T as 2000 sec., the value of this velocity is about 0·03 cm. sec^{-1} and will, of course, be greater for the higher frequency terms in the Fourier series of simple harmonic terms which represents the actual applied temperature wave. This means that the temperature wave will have travelled the thickness of the steel plates in a time of the order of 100 sec. The damping factor $\sqrt{(\pi/hT)}$ is small in this case, being approximately 0·1, so that inside the steel at 1 cm. from the liquid/steel boundary the temperature excess is 9/10 of the value at the interface. Hence, during the time of flow with the electric field applied, the whole of the metal, of thermal capacity 260 cal. deg^{-1} c., may be considered to a first approximation as being at the temperature of the liquid. The heat which must be given to the liquid and the surrounding metal in order to produce the rise of temperature noted (0·04° c.) is thus 10·8 calories.

During the flow the whole surface of the apparatus is losing heat to the air bath, whose temperature is 0·04° c. lower. Taking the surface area as 320 cm^2 and its emissivity as $2{\cdot}5\times10^{-4}$ cal. $cm^{-2}\ sec^{-1}\ deg^{-1}$ c., the heat given to the enclosure is about 3·2 calories, so that the total quantity of heat generated in the liquid is 14·0 calories in 1015 sec. Since the volume of liquid present at any instant in the capillary is 0·143 cm^3, the mean rate of production of heat in unit volume of liquid is $9{\cdot}6\times10^{-2}$ cal. $cm^{-3}\ sec^{-1}$.

Theoretically the condenser with dielectric may be considered as made up of a pure capacity in series with a small pure resistance. The rate of dissipation of energy W, per unit volume of dielectric, when an alternating field of amplitude X and frequency ν is applied to the system is then $\frac{1}{4}X^2K\nu\cos\phi$, where K is the dielectric constant and $\cos\phi$ is the power factor. In the case considered, $X=35$ kv. $cm^{-1}=1{\cdot}17\times10^3$ E.S.U. cm^{-1}, $\nu=10^4\ sec^{-1}$ and $K=6$, so that $W=4{\cdot}9\cos\phi$ cal. $cm^{-3}\ sec^{-1}$.

Equating this expression for W to that obtained by experiment gives the value 2×10^{-2} for the power factor for chlorbenzene under the conditions holding in the experiment.

Similar effects have been observed with higher fields and frequencies for monochlorbenzene, chloroform and amyl acetate, all of which are polar.

No change in time of flow greater than 1 part in 100,000 was observed with benzene. This liquid is non-polar, and any dielectric heating can only be obtained through the distortion polarization of the molecule, and would, therefore, be expected to be very small, since the power factor is, in this case, of the order of 2×10^{-4}.

Physics Dept., University College, London. C. Dodd.
20th April 1948.

Andrade, E. N. da C., and Dodd, C., 1946, *Proc. Roy. Soc.* A, **187**, 299.

The B-N Band-system of Silver Iodide

In the course of work on the absorption spectrum of silver-iodide vapour in the far ultra-violet region (Barrow and Mulcahy 1948) it was noticed that within a rather narrow range of pressure the centre of the B–N system at 3200 A. was well developed. The long wavelength end of this system has been the subject of careful investigation by Brice (1931), but his measurements end with the 0,4 band at 3295 A., and in the region 3170–3295 A. only the rather rough data of Franck and Kuhn (1927) are available. We have therefore thought it useful to record our measurements, especially since they lead to somewhat different values of ω_e' and D' from those given by Brice.

The bands were photographed in a first order of a 2·4-m. grating instrument on Ilford Process plates, the linear dispersion on the plate being about 7·4 A/mm. A tungsten filament lamp with " Cornex " envelope, kindly supplied by the Research Laboratories of the G.E.C., was used as source of continuum. The optimum temperature of silver iodide was about 900° C. Of the sixteen bands measured, only for the 2,4 and 0,3 bands was the silver isotope separation clearly resolved : for the other bands it was assumed that settings on the low-λ side of the origin were on ^{107}AgI heads and on the high-λ side on ^{109}AgI heads.

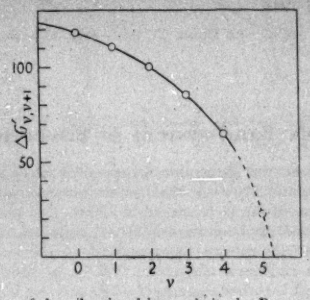
Table 1. Band-head data for ^{107}AgI

Italic figures denote wavelengths in air; roman, wave numbers *in vacuo*; figures in parentheses are visual relative intensities.

v' ↑	v'' → 0	1	2	3	4
6	31657 (0) *?3157·9* (23)				
5	31633·6 (1) *3160·28* 64·9				
4	31568·7 (2) *3166·78* 86·4				
3	31482·3 (3) *3175·47* 98·2			30868·4 (1) *3238·63* 99·6	
2	31384·1 (7) *3185·41* 111·3	206·9 31177·2 (3) *3206·55* 110·1		30768·8 (2) *3249·11*	200·8 30568·0 (2) *3270·45*
1	31272·8 (6) *3196·75* 118·4	205·7 31067·1 (7) *3217·91* 119·0	205·8 30861·3 (3) *3239·37* 116·7		
0	31154·4 (5) *3208·90*	206·3 30948·1 (9) *3230·29*	203·5 30744·6 (10) *3251·67*	204·2 30540·4 (8) *3273·41*	

A correction was applied to the latter, so that all the figures in table 1, which summarizes the band-head data, relate to ^{107}AgI.

The results, particularly for the $v''=0$ progression, provide new information about the vibrational levels of the excited state. This is given in table 2 in the form of energies of levels v' in excess of $v''=0$ (for ^{107}AgI), calculated both from the bands in table 1 and from Brice's figures for ^{107}AgI and ^{109}AgI. In the calculations, Brice's expression for G''_v has been used. The agreement is very satisfactory, and shows that the expression for G''_v deduced for bands with v'' ranging from 4 to 17 holds well over the range 0 to 17. From the mean results of table 2, values of $\Delta G''_{v,v+1}$ were determined and plotted as in the figure.



Course of the vibrational intervals in the B state of ^{107}AgI.

We obtain $\omega'_0=123{\cdot}5$, $D'_e=570$, $D'_0=510$, which differ significantly from the figures given by Brice, namely, $\omega'_0=131{\cdot}3$, $D'_e=769$ cm^{-1}.

Table 2. Energies of levels v' in excess of $v''=0$ (cm^{-1})

v'	Present work	Brice	Mean	$\Delta G'$
6	31657) (1)*	—	(31657)	
5	31633·6 (1)	—	31633·6	(23)
4	568·7 (1)	31568·6 (4)*	568·6	65·0
3	482·0 (2)	482·9 (14)	482·8	85·8
2	383·3 (5)	382·3 (12)	382·6	100·2
1	272·1 (3)	272·0 (8)	272·0	110·6
0	153·9 (5)	153·9 (3)	153·9	118·1

* Number of band-heads from which the values were derived.

The unusual character of the B state (Mulliken 1937) is emphasized by these results. Thus, from the distribution of intensity among the bands, it appears that r'_e is somewhat, but not very much, greater than r''_e; yet the force-constant in state B is only 36% of its value in the ground state, while the vibrational levels converge so rapidly that D'_0 is only 3% of the dissociation energy of the ground state.

Physical Chemistry Laboratory,
Oxford.
22nd April 1948.

R. F. Barrow.
M. F. R. Mulcahy.

Barrow, R. F., and Mulcahy, M. F. R., 1948, *Nature, Lond.*, in the press.
Brice, B. A., 1931, *Phys. Rev.*, **38**, 658.
Franck, J., and Kuhn, H., 1927, *Z. Phys.*, **43**, 164.
Mulliken, R. S., 1937, *Phys. Rev.*, **51**, 310.

DISCUSSION

on paper by R. Weil entitled "The Variation with Temperature of Metallic Reflectivity" (*Proc. Phys. Soc.*, 1948, **60**, 8).

Dr. D. J. Price. The equation preceding (15) has the obvious factor $2x-a$ leading to a root having a value $\frac{1}{2}a$. This root leads to λ_x being of the magnitude $0{\cdot}1\ \mu$ for most metals. If, as stated, the cubic had only one real root, the above would be the only solution. This, however, is also false. It is probable that the complicated substitutions and expansions have given stationary points for that portion of the reflectivity curve giving negative values for the reflectivity. (These are, of course, without physical significance.)

From purely physical grounds, the conclusion of (16) may be seen to be unsound, since λ_r is the limit at which the Hagen-Rubens relationship becomes valid to a high approximation, and, therefore, the anomalous behaviour associated with the appearance of an X-point cannot occur here.

Taking all this into account, it would appear then that the only region in which an X-point may be found is in wavelengths small enough for the contributions of the bound electrons to be considerable. Recent work, to be published shortly, shows that this is indeed the case, and that a metal containing free electrons only can have no X-point.

Author's reply. I am very grateful to Dr. Price for the comments he has made in his discussion. It is true that one solution of my equation (15) is $x=a/2$. As Dr. Price himself points out, this gives rise to an X-point at $\lambda=0{\cdot}1\ \mu$; this wavelength is so far in the ultra-violet region that my original assumption of free electrons being responsible for the mechanism of reflection cannot be maintained. Therefore this solution can be rejected. On dividing (15) by $2x-a$ there remains the quadratic $2x^2-(3a-2y)x+ay=0$, which gives

$$X=\tfrac{1}{4}\{3a-2y\pm\sqrt{(9a^2-20ay+4y^2)}\}\simeq\tfrac{1}{4}\left[3a-2y\pm3a\left(1-\frac{10}{9}\frac{y}{a}+\frac{2y^2}{9a^2}\ldots\right)\right]. \quad \ldots\ (1)$$

The positive sign gives an X-point which falls still into the region where the effect of bound electrons predominates, and only the negative sign yields a value comparable with those obtained experimentally by Dr. Price. Since $y/a\simeq1/700$ it follows from (1) above that $x=\frac{1}{4}(-2y+10y/3\ldots.)=y/3$, which yields my equation (16). Since (16) has been tested for nickel by computing values for λ_r from katoptric data, and the λ_x predicted by this equation is confirmed experimentally, I fail to see why the conclusions drawn from this formula should be unsound. Nor can I understand how Dr. Price arrives at the conclusion that stationary points in $dR/d\nu$ presuppose negative values for the reflectivity. However, experiments are in progress which are designed to test the temperature variation of λ_x; results will be reported on completion of this work.

DISCUSSION

on paper by G. F. J. Garlick and A. F. Gibson entitled "The Electron Trap Mechanism of Luminescence in Sulphide and Silicate Phosphors" (*Proc. Phys. Soc.*, 1948, **60**, 574).

Dr. H. A. Klasens. Lord, Rees and Wise (*Proc. Phys. Soc.*, 1947, **59**, 473), have explained their decay results also by assuming the electrons to be trapped near excited centres, but they allow for the possibility of such electrons escaping to the conduction band and combining with other excited centres. Does Dr. Garlick allow for this process? Also, how does he explain that Cu introduced into zinc sulphide activated with Mn prolongs the afterglow of the Mn emission?

AUTHORS' reply. The assumptions made by Lord, Rees and Wise in their paper are not necessary to explain our experimental results. They are based on very restricted phosphorescence decay measurements on one specimen of zinc sulphide activated by zinc and silver with a trace of copper. No independent study of the trapping states was made using thermoluminescence measurements. It would be difficult to distinguish the effects of certain traps being related to different emission centres in such a phosphor since the emission spectra of the two different centres formed by zinc and silver are almost identical.

With respect to the effect of copper in prolonging the afterglow of zinc sulphide phosphors activated by manganese, it must be assumed that electrons from traps associated with the copper impurity cause emission in manganese centres when they escape. However, it is not essential that the transit between centres should take place via the conduction levels of the matrix. Such a process, involving entry of electrons into the conduction levels, would be temperature dependent, whereas in some specimens of this type of phosphor we have found no marked temperature effect. Since electron traps have a large effective size (Garlick and Gibson, *Proc. Roy. Soc.* A, 1947, **188,** 485), "overlap" of trapping states belonging to different centres is likely which may allow electron transfers. This "overlap" of traps can also explain the difference in the copper concentrations for optimum afterglow and optimum emission efficiency in zinc sulphide phosphors.

REVIEWS OF BOOKS

Science and the Meanings of Truth, by MARTIN JOHNSON. Pp. 179. (London: Faber & Faber, 1946.) 12*s.* 6*d.* net.

This book continues the line of thought which Dr. Johnson has followed in two earlier works, the emphasis here being laid on the meanings of "truth" in modern physics. (Somewhat unfortunately the words "truth" and "truthfulness" are used indiscriminately: the reader who is accustomed to understanding by the latter the laudable but limited quality exhibited by the youthful George Washington must learn to give it a wider meaning here.) In the older physics causal explanations of phenomena in terms of inferred entities resembling phenomena were sought through hypotheses, and a hypothesis was true if it explained the facts without exception. In modern physics this aim is abandoned and it is recognized that physics reveals only a structure characterizing entities about which we know nothing. The word "truth" is thus left without a meaning, and the author suggests that it should be identified with "communicability" and "coherence" as characteristics of the structure. The arguments for this view occupy the major part of the book. The remainder is concerned with the philosophical problem of the relation between the structure and our sense experience and finally with the relation between the scientific and non-scientific realms of experience. For reasons of space, only the first part, which most directly affects physicists, will be discussed in this review.

The book reveals, as would be expected, an intimate knowledge of the present situation in physics and allied departments of thought. It is not easy reading but it repays the necessary effort. The fundamental problems aroused by relativity and quantum theory are considered in relation to the modern branch of logic known as "logistics", especially the calculus of propositional functions, but strangely enough the post-logistics developments, such as those of the metamathematicians and the intuitionists, which transcend logistics almost as thoroughly as that transcended scholastic logic and are much more vital for the subject treated, are not mentioned. On the question whether the laws of nature are empirical or rational in origin, Dr. Johnson's view seems to be that they are both. As a disciple of Milne he thinks that they can be derived from "our intuitive appreciation of the passage of time", which is taken as empirical, without further observation, by the use of reason alone. At least, I think that this is his view.

The core of the book, however, is the thesis that "communicability" is the dominating criterion of truth in physics. Dr. Johnson has emphasized the importance of communicability in each of his previous works, and it is evidently invested in his mind with profound significance. Unfortunately, however, he uses the word in so many senses that it ceases to

have any consistent meaning. In one place a thing is communicable if it can be communicated from one observer to another. That is legitimate, but shortly afterwards it is implied that the two observers must *agree* about what is communicated, and that is not at all implied in the proper use of the word. Yet again, a scientific proposition, to be true, must be " communicable from one experimental situation to another ", by which is meant that a conception, to be true, must be applicable to different, originally independent, fields of investigation. For example, whenever a wave-like property, such as a measurable frequency or a diffraction pattern, appears in a communicable set of experiments, then this particular set achieves communication in turn with other situations which have been characterized by wave-like properties. This state of affairs is also called " coherence ". If this variation of meaning merely caused confusion it would be a remediable nuisance, but it is in fact a fatal flaw because a situation, having been endowed with communicability in one sense, is thereupon assumed to possess it in others, and so a fallacy is introduced. It is easy to prove that black is white if one is allowed to call them both grey.

The most striking instance occurs in the treatment of the Lorentz transformation. It is a fact that this transformation makes different observations (such as those of the orbital motion of the Earth and the non-displacement of fringes in the Michelson-Morley experiment) coherent. Hence it establishes " communicability " and so is assumed to demand the existence of a host of observers to make the communications which enable them to agree with one another. All this is simply false. The Lorentz transformation has nothing to do with different observers ; it concerns the way in which a single observer changes his coordinate system. This is not opinion, it is obvious fact. Lorentz did not derive the transformation so that he could agree with the Man in the Moon. He derived it so that different observations by what was effectively one observer—an earthbound one—could be reconciled. Hence, since virtually only one observer is concerned, the question of " communicability " in the proper sense of the word does not arise. The position can perhaps be seen more clearly in relation to a simpler change of coordinate systems—one in which we merely change the origin of coordinates. The statement that two places are respectively 10° and 15° east of Greenwich is rejected as a fundamental statement not because it is not communicable ; it is communicable and, furthermore, both parties to the communication can agree about it, and the statement preserves the coherence of all geography. But it is rejected because Greenwich is an arbitrary starting-point and the simpler statement that the places differ in longitude by 5° is preferred because it is free from such arbitrariness. That has nothing to do with different observers ; it holds good if there is only one observer in the world. With the Lorentz transformation the position is exactly the same, with velocities substituted for positions. That also has nothing to do with a multitude of observers ; it is concerned with the various coordinate systems that a single observer may choose. The illusion that different observers, and communicability, are involved is the result of the most unfortunate device that early expositors adopted in order to explain the theory of relativity to the man in the street. He was assumed not to be familiar with coordinate systems but to have a natural tendency to think himself at rest (no man of science has done so since shortly after the time of Copernicus). The second coordinate system was therefore personified into a second observer who also thought himself at rest.

All this, I repeat, is not opinion but simple, undeniable fact. As in his earlier works, Dr. Johnson overlooks it and bases his argument on the supposed actuality of imaginary observers. He is not a dishonest thinker, and one can only conclude that he is so charmed by Milne's theory of kinematical relativity, his devotion to which is manifest throughout his writings, that he cannot bring himself to discard the fictitious situation on which Milne himself bases it. For if—acknowledging as he does, the Lorentz transformation to be " the classic instance " of what he calls the " attainment of communicability "—he takes that transformation in its legitimate, unperverted form, he will find that he cannot proceed to the conclusion he so much desires, and the whole theory of kinematical relativity will be exposed as a baseless structure. " Communicability " makes a host of observers a necessity for the *possibility of existence*, not merely the practicability of prosecution, of science, and this is fundamentally false. A single mind, with unlimited intelligence, mechanical skill and freedom of movement, and sufficient time, could derive and verify the whole of our physical knowledge without anything with which it could " communicate ". If any one doubts this he has only to try to name a piece of knowledge not so obtainable. Any philosophy which is inconsistent with this fact stands self-condemned.

What makes this characteristic of Dr. Johnson's work the more regrettable is that he is one of the few writers who, in this age of specialization, have the breadth of vision and the depth of thought necessary to grasp the meaning of the scientific movement as a whole and estimate its significance. He has something to say and has the ability to say it. It is profoundly to be hoped that he will have the courage to discard fanciful speculations and reform his contribution to the philosophy of our time so that it can rest on foundations true in the George Washington sense of the word. HERBERT DINGLE.

Magnetic Materials, by F. BRAILSFORD. Pp. ix+156. (London: Methuen & Co. Ltd., 1948.) 6*s*.

Dr. Brailsford, as a member of the Research Department of the Metropolitan-Vickers Electrical Company, has had ample opportunity of first-hand acquaintance with the many new magnetic materials which have found, and are still finding, their way into all kinds of laboratories and workshops. It is natural that he should open his survey of these materials and their properties with a short, but adequate, introduction to modern theories of ferromagnetism, the properties of ferromagnetic single crystals and the various factors which determine the magnetic behaviour of metals and their alloys. This introduction, which takes up more than half of the booklet, is excellently illustrated with clear diagrams.

The materials themselves are studied under three headings: Iron and Silicon-Iron Alloys, Nickel-Iron and other Alloys, and Permanent Magnet Materials. All students should find these sections very helpful. Lists of original references are given, and it should therefore be easy to follow up particular points of interest. The author has certainly packed a great deal of useful information into a very small space, mainly by the wise use of graphs. The book can be unreservedly recommended to all interested in the subject.

The reviewer was struck by the use of the term *retentivity* on page 135 to denote *residual inductance* or *remanence*, for which Ewing used the term *retentiveness*. As far as can be ascertained, the word retentivity has hitherto been used to denote residual intensity of magnetization. L. F. BATES.

Précis de Physique Générale, II: La Chaleur, by P. and A. MERCIER. Pp. 105. (Neuchâtel: Editions du Griffon, 1947.) 12 fr. (Swiss).

One of the authors is an engineer, former dean of the scientific faculty at the Collège de Genève, and the other is professor of theoretical physics at Berne. Their joint work should, therefore, present a good balance between the theoretical and the practical outlooks. They have, in fact, written a very good summary of the fundamentals of classical heat theory. Classical thermodynamics is presented in a very sound manner, with a very clear treatment of the basis of the second law and the definition of absolute temperature, but there is no reference to modern quantum mechanics and none to statistical mechanics. Of the subjects dealt with, it is interesting to note that whilst internal energy and entropy are introduced, the only functions derived from them is Gibbs' function, here called the *thermodynamic potential*. No mention is made of the function known as heat content or enthalpy, nor of Helmholtz's free energy.

The book is thus very good for those who desire a brief development of thermodynamics, with the practical side of thermometry and calorimetry, but it needs supplementing for such matters as heat conduction and radiation. J. H. A.